DOCUMENT RESUME

ED 353 559 CS 011 161

AUTHOR Bruce, Bertram C.; And Others
TITLE The Content and Curricular Validity of the 1992 National Assessment of Educational Progress in Reading.
INSTITUTION Center for the Study of Reading, Urbana, IL.
SPONS AGENCY National Academy of Education, Washington, D.C.
REPORT NO CSR-TR-569
PUB DATE Feb 93
NOTE 56p.
PUB TYPE Reports - Descriptive (141) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Content Validity; Elementary Secondary Education; *Reading Achievement; Reading Research; *Reading Tests; *Student Evaluation; *Test Construction; *Test Content
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

One of 10 reports commissioned by the National Academy of Education, this report investigates topics related to an assessment program piloted by the National Assessment of Educational Progress (NAEP) designed to support state-by-state and state-to-nation comparisons of student performance in reading. The report focuses on three issues: (1) the adequacy of the process used to develop the "Reading Framework for the 1992 NAEP in Reading"; (2) the degree to which the "Framework" represents a consensus about reading among researchers, practitioners, and state and local school administrators; and (3) the extent to which the fourth-, eighth-, and twelfth-grade levels of the assessment exemplify the recommendations of both the "Framework" and the document written to guide the selection of passages and the development of assessment items. The report begins with a review of some of the events that led to the development of both the "Framework" and the 1992 NAEP in reading. Following an overview of the "Framework," the report discusses the three issues, describing the methods used to gather information and findings. The report next looks at the special studies that were part of the assessment. Finally, the report offers recommendations. Four tables of data are included; the survey cover letter, a list of the interview questions and a copy of the questionnaire (with percentages of responses for each item) are attached. (Author/RS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED353559

Technical Report No. 569

THE CONTENT AND CURRICULAR VALIDITY OF THE 1992 NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS IN READING

Bertram C. Bruce
Jean Osborn
University of Illinois at Urbana-Champaign

Michelle Commeyras
University of Georgia

February 1993

Center for the Study of Reading

TECHNICAL REPORTS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Fran Lehr

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

College of Education
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
174 Children's Research Center
51 Gerty Drive
Champaign, Illinois 61820

CS011171

CENTER FOR THE STUDY OF READING

Technical Report No. 569

THE CONTENT AND CURRICULAR VALIDITY OF THE 1992 NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS IN READING

Bertram C. Bruce
Jean Osborn
University of Illinois at Urbana-Champaign

Michelle Commeyras
University of Georgia

February 1993

College of Education
University of Illinois at Urbana-Champaign
174 Children's Research Center
51 Gerty Drive
Champaign, Illinois 61820

1992-93 Editorial Advisory Board

Diane Bottomley

Eurydice M. Bouchereau

Clark A. Chinn

Judith Davidson

Colleen P. Gilrane

Heriberto Godina

Richard Henne

Carole Janisch

Christopher Kolár

Brian A. Levine

Elizabeth MacDonell

Montserrat Mir

Punyashloke Mishra

Jane Montes

Billie Jo Rylance

Shobha Sinha

Melinda A. Wright

MANAGING EDITOR
Fran Lehr

MANUSCRIPT PRODUCTION ASSISTANT
Delores Plowman

Abstract

In 1992, the National Assessment of Educational Progress (NAEP) piloted an assessment program designed to support state-by-state and state-to-nation comparisons of student performance in reading. This report is one of 10 commissioned by the National Academy of Education to investigate topics related to the NAEP. It focuses on three issues: (a) the adequacy of the process used to develop the *Reading Framework for the 1992 NAEP in Reading*; (b) the degree to which the *Framework* represents a consensus about reading among researchers, practitioners, and state and local school administrators; and (c) the extent to which the fourth-, eighth-, and twelfth-grade levels of the assessment exemplify the recommendations of both the *Framework* and the document written to guide the selection of passages and the development of assessment items. Interviews, panel meetings, colloquia discussions, and a survey questionnaire were used to address these issues. The report begins with a review of some of the events that led to the development of both the *Framework* and the 1992 NAEP in Reading. Following an overview of the *Framework*, the report discusses the three issues, describing the methods used to gather information and findings. The report next looks at the special studies that were part of the assessment. Finally, it offers recommendations. Appendices contain a list of the interview questions and a copy of the questionnaire, with percentages of responses for each item.

THE CONTENT AND CURRICULAR VALIDITY OF THE 1992 NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS IN READING

In 1990, the National Assessment of Educational Progress (NAEP) piloted an assessment program designed to collect data from individual states that would permit state-by-state and state-to-nation comparisons of student performance in mathematics. In 1992, NAEP conducted a similar pilot assessment in reading. This report describes one of 10 studies commissioned by the National Academy of Education to investigate topics related to those two pilot assessments.

In light of the importance of the first state-by-state reporting of NAEP reading data in 1992, the National Academy asked us to undertake an investigation of three issues related to the 1992 NAEP Assessment in Reading: (a) the adequacy of the process used to develop the *Reading Framework for the 1992 National Assessment of Educational Progress*, the document written to provide a foundation for the development of the assessment; (b) the degree to which the *Framework* represents a consensus about reading among researchers, practitioners, and state and local school administrators; and (c) the extent to which the fourth-, eighth-, and twelfth-grade levels of the assessment exemplify the recommendations of both the *Framework* and the *Assessment and Exercise Specifications*, the document written to guide the selection of passages and development of items.

The National Academy then asked us to consider four additional questions. The first has to do with the field of reading, the second with the nature of the students in the nation's schools, and the third and fourth with the assessment itself:

1. How can the assessment results best be presented to the professionals in the field of reading, given the fact that there are no clearly defined and agreed-upon guidelines for the teaching of reading?

2. Does the proposed assessment adequately address the issues of linguistic diversity and varying background knowledge of a multicultural student population?

3. Given that it is common practice to adjust teaching so that students will do well on tests, how will student performance be affected by the implementation of the assessment?

4. How will the results of the assessment be explained to the public and policy makers, given the possibility that large numbers of students may do poorly on it?

We begin our report with a review of some of the events that led to the development of both the *Framework* and the 1992 NAEP in Reading. We believe this information is important to an understanding of why the National Academy focused on the particular topics and questions listed above. We then give an overview of the *Framework*. Following the overview, we address the three topics we were asked to examine by the National Academy. We describe the methods used to gather information about each topic and discuss our findings. Finally, we summarize our responses to the four questions.

When we began this project, we were well aware of the sharp divisions within the field of reading over a number of issues such as beginning reading instruction, type of instruction, ability grouping, round-robin reading, and how reading should be assessed. Thus, a key question for us was whether we could judge the match of the *Framework* to the consensus of the field when there apparently was so little consensus.

As we gathered data, we confirmed that divisions indeed do exist within the reading field. But to our surprise, we also found a remarkable consensus about the strengths of the *Framework*. We reached this conclusion through our examination of the *Framework*; by listening to the discussions among the members of the panel we convened and to those occurring at the two colloquia we held at the Center for the Study of Reading; in interviewing over 50 leaders in reading; and in reading the responses to the survey questionnaire we developed and sent to educational leaders throughout the United States.

Just as there was general agreement regarding the strengths of the *Framework*, there were also concerns. But there was much less consensus about the concerns than about the strengths. We believe, however, that the concerns raised by participants in our initial investigations are worthy of consideration in this report.

The National Academy defined a limited focus for this study. This focus excluded a number of questions that are central to evaluating the NAEP in Reading. Among these are the following: (a) Is a national assessment a good approach overall? (b) What are the political and social implications of state-by-state reporting of the assessment data? (c) Should a large-scale assessment be constructed from the "top" or be built up from assessments situated in classroom practice?

Many in the field of reading argue that extensive knowledge of educational attainment already exists in the experiences of teachers and other practitioners, and that, moreover, this knowledge is already situated with respect to the cultural, social, and institutional contexts in which students learn. This knowledge is based on longitudinal observations of students' performance on a variety of tasks, including collaborative and cross-disciplinary work. Thus, a radical alternative to the NAEP approach would be to look for ways in which this knowledge could be made more widely known to meet the needs of other teachers, parents, citizens, and policy makers.

In our study, we did not address questions and points of view such as these, rather we acknowledged that there would be a national, large-scale assessment of reading in 1992 and that state-level data would be reported for those states that volunteered to be included in the trial state assessment.

The question we did address is: "Within the paradigm adopted by the federal government, how well do the *Framework*, the passages, the items of the assessment, and the scoring accord with the views of experts in the field of reading?"

Background Information: Preparation for the 1992 NAEP in Reading

In 1989, the National Assessment Governing Board (NAGB) awarded the Council of Chief State School Officers (CCSSO) a one-year contract to organize the NAEP Reading Consensus Project. This project was to develop a set of five documents, each of which was to consider an aspect of reading or its assessment that was relevant to the development of the 1992 NAEP in Reading. These documents were to be written by project staff at the CCSSO with the advice of members of NAGB, the National Center for Education Statistics (NCES), and two committees appointed to work on the project.

In this report, our primary focus is on one of these documents, the *Reading Framework for the 1992 National Assessment of Educational Progress*, which provides the rationale for the form and content of the assessment. Our secondary focus is on another document, the *Assessment and Exercise Specifications*, which contains instructions for the selection of passages and the development of assessment items. We also utilize the Reading Consensus Project's final report, *Report of the Consensus Process*. The *Framework* is addressed to reading professionals and to members of the public interested in the approach to reading and assessment that undergirds the assessment. An equally important audience for both the *Framework* and the *Specifications* was the Educational Testing Service (ETS), the contractor for the development of the assessment.

BEST COPY AVAILABLE

In awarding the CCSSO the contract, NAGB advised that an assessment be developed that would, as much as possible, reflect a consensus of the views of the people in the field of reading. In addition, it was intended that the assessment should give evidence of being *modern*, in that it would reflect current research and knowledge about the reading process and the assessment of reading competence.

The difficulties associated with achieving these goals must not be minimized. Reading educators, researchers, and others in the field are well known for their diverse and often conflicting opinions about the nature of the reading process, appropriate approaches to reading instruction, and meaningful assessment of reading competence. The task of the Reading Consensus Project was to provide a forum for the development of an assessment that would represent the most agreed-upon views of reading and its assessment. NAGB charged this group with the job of making the *Framework* a document that would make explicit the rationale for the both the form and content of the 1992 NAEP in Reading, and would, as the document was developed, attempt to include the views, opinions, and reactions of a number of reading researchers and educators.

A radical change in how and to whom NAEP scores would be reported was the primary reason for devoting additional attention to the content and quality of the assessment. Because the decision had been made to report (on a trial basis and with only the fourth-grade scores) the results of the assessment on a state-by-state basis, the need for establishing some consensus within the field of reading about the content and form of the assessment was intensified.

In previous years, NAEP findings were always reported as national-level data. Since 1969, seven national-level NAEPs in reading have been conducted. In fact, these assessments represent the only continuing assessment of reading achievement in the United States. The information from these assessments has been used to compare nationally representative groups of fourth-, eighth-, and twelfth-grade students on the basis of ethnicity, gender, and the type of community and the region in which the students live. In addition, NAEP reporting has included trend data that reflect changes in student performance over time, as well as some data that correlate reading achievement and such student activities as time spent on homework. In recent years, these reports have been released to the public as *The Nation's Report Card*.

During the past decade, a number of state and national educational and political leaders have expressed interest in the state-by-state reporting of NAEP results. In 1984, a majority of chief state school officers supported the development of an assessment that would permit state-by-state reporting. In 1985, this group suggested that the NAEP would be "the most feasible vehicle for such an assessment." Two years later, a group appointed by Secretary of Education William Bennett recommended that the assessments be extended to provide for state-by-state reporting. Subsequently, 37 states volunteered to participate in the 1990 trial state assessment for mathematics. About 45 states volunteered for the 1992 trial state assessments, which include the fourth-grade reading trial.

It must be noted that a number of educators have expressed concerns about the wisdom of reporting NAEP data at the state level. These include, for example, concerns that: (a) the results of assessments will be used to draw inappropriate conclusions about student performance, which in turn may inadvertently lead to damaging policy decisions; (b) the content and form of the assessment may not match the goals of public education; and that (c) any attempt at large-scale assessment will fail to capture the complexity of its subject matter and thus provide an inappropriate model of instruction for the teachers and students in American schools.

Knowing the significance of the decision to do state-by-state reporting, and being aware both of the lack of agreement within the field of reading on a number of issues and of the concerns about state-level reporting, the Reading Consensus Project began its work in October 1989. Two committees were appointed: a Steering Committee, composed of representatives from a number of professional

educational organizations and the National Alliance for Business, and a Planning Committee, composed primarily of experts in reading research, the implementation of reading instruction, and the assessment of reading. A staff member from the Chase Manhattan Bank represented the public on this committee.

One task of the Steering Committee was to identify basic principles and policies for the Planning Committee and to respond to that committee's progress reports. The task of the Planning Committee was to set content objectives and technical features for the assessment and to reach -- among its members, and in consultation with a wide variety of people in the reading field -- a consensus about the content and form of the assessment. It should also be noted that this committee's decisions about the content and form of the assessment had to be communicated to and be approved by the Steering Committee and NAGB. The Reading Consensus Project's final report summarizes the roles of these two committees: The Steering Committee set a position based on broader social policy needs, and the Planning Committee represented values involved in reading.

In October 1989, the Steering Committee prepared a set of guidelines to frame the work of the Planning Committee. These guidelines asked for a reading framework that would

- Focus on results, rather than on specific methods for teaching reading;
- Be real-world oriented by addressing the nation's changing literacy needs for employability, personal development, and effective citizenship;
- Be innovative, by supporting the expansion of existing assessment strategies to include more open-ended questions, non-traditional approaches, and new formats;
- Respond to the latest scholarship on reading theory and instruction;
- Create information for policy makers that can help support informed decisions;
- Provide a forum for discussion of what is reasonable for students to know and be able to do as they read.

The Planning Committee met five times during the months of November and December. Because of the short time line, the group was pulled together hastily, making it impossible for all members to attend each of the meetings. The project staff worked hard, however, to communicate the proceedings of each of the meetings to members not present. These members were kept up to date, both by mail and through conference calls. The Project Coordinator was diligent about recording all comments, both oral and written, that were provided her in response to these communications.

It should be noted that a number of observers attended the meetings. Among these were representatives of the Steering Committee, NAGB, ETS, and NCES. In addition, reports of the meetings were sent to a number of other people in the reading field who were not on the committee.

Throughout their deliberations, the committee members and the Reading Consensus Project staff were constantly aware of the criterion (set forth for the *Framework* by NAGB) that it be accessible to the interested public as well as credible to members of the reading community. By the end of January 1990, the first draft of the *Reading Framework for the 1992 National Assessment of Educational Progress* was complete.

As the Planning Committee developed the *Framework*, efforts were also underway to develop assessment specifications. This work was with the help of the American Institute for Research (AIR), Palo Alto. In the initial stages of this project, several members of the Planning Committee worked with staff

members at AIR. The *Assessment and Exercise Specifications* report was completed at approximately the same time as the *Framework*, and the two documents were forwarded to the Steering Committee and to ETS.

Because the *Framework* provides the rationale and recommendations for the assessment, and because it is the document that is central to this report, its content is reported in considerable detail in the next section.

The Reading Framework for the 1992 NAEP

The information in the *Framework* is organized into introductory material and six major sections entitled (a) Guiding Considerations, (b) 1992 Reading Literacy Objectives, (c) Types of Text, (d) Cognitive Aspects of Reading, (e) Constructing the Assessment, and (f) Special Studies. An appendix contains sample passages and items and lists of members of various committees.

Introductory Material

The introduction to the *Framework* begins with a brief review of previous NAEP assessments in reading and of the events leading to the decision to report NAEP data to individual states. It then identifies the factors used to guide the development of the 1992 assessment. These factors are as follows: (a) The general pattern of consensus development, set forth by law and evolving over time, calls for "active participation of teachers, curriculum specialists, subject matter specialists, local school administrators, parents, and members of the general public"; (b) the fact that the assessment will pilot state-by-state comparisons, which increases the importance of the consensus process; (c) the recognition of the diverse and conflicting views of reading "that have not been completely illuminated, much less settled, by research"; and (d) a time frame for the process that is shorter than ever before, while the stakes are higher.

The section concludes with descriptions of the duties of the Steering and Planning Committees of the Reading Consensus Project, the major events of the development process, and the list of guidelines the Steering Committee presented to the Planning Committee at their first meeting.

Guiding Considerations

The main body of the *Framework* opens with statements about the considerations and principles that governed its development. A condensed version of these follows.

1. The NAEP in Reading is an *assessment*, not a test. Assessments are designed to provide information about progress or achievement in general rather to test ability relative to a predetermined standard. The NAEP is designed to inform policymakers and the public of the state of reading in the United States in broad terms.

2. The NAEP uses the term "reading literacy" to connote a broader sense of knowing when to read and how to read and reflecting on what we read afterward. It is *not* intended to mean basic or functional literacy.

3. Assessment by itself should not drive instruction. One goal of the NAEP is that its content be valid and authentic so that it would be appropriate for teachers to teach toward the areas it suggests. Another goal is that it be so broad and complete in its coverage of important reading behaviors that it would still be valid, useful, and appropriate if teachers or schools consciously addressed the kinds of things it covers.

4. The facets of reading that can be measured in a project of national scope are limited at this time. Therefore, the best use must be made of available methodology and resources and efforts must be undertaken to improve measurement techniques.

5. The legislation for NAEP authorized that the 1992 assessment in Reading increases concern about the strength of the assessment design and about how results will be reported. Aware of these concerns and the controversy surrounding the state-by-state assessment, the Planning Committee must make every effort to consider a variety of opinions, perspectives, and emphases among professionals and state and local school districts in developing the *Framework*.

1992 Reading Literacy Objectives

Asserting that the goal of literacy education is to develop good readers, this section of the *Framework* begins with a listing of the characteristics that identify good readers: (a) Good readers exhibit positive reading habits and value reading; (b) they read with enough fluency so that they can focus on the meaning of what they read, rather than devoting a lot of attention to puzzling out words; (c) they use what they already know to understand the text they are reading -- they extend, elaborate, and critically judge the meaning of the text; and (d) they plan, manage, and check the progress of their reading and use effective strategies to aid their understanding.

The *Framework* proposes these characteristics, verified by research and experience, as the guide to what should be assessed in reading. It backs this proposal with the statement that "the orientation toward good readers reflects a focus on performance as an end product rather than a focus on instructional approaches in reading."

The *Framework* defines reading as a constructive, dynamic process, rather than as a collection of related subskills: "Reading is a deep, specific interaction between the reader, the text, and the situation." It also highlights the importance of prior knowledge, as well as "a degree of understanding and skill in reading," and acknowledges that a reader's way of reading changes in response to the purposes for reading and to the type of text being read.

Types of Text

Within the *Framework*, the two sections, Types of Text and Cognitive Aspects of Reading, describe the most important features of the assessment. These sections define the kinds of texts to be used in the assessment, the expanded view of reading that is the basis of the *Framework*, and the rationale for the construction of items.

The *Framework* points out that, "depending on the text itself and the reader's purpose for reading, the reader is oriented to a text very differently." It proposes that because of the differences in reading behavior that result from reading various texts for a variety of purposes, the assessment should contain three broad categories of text. These categories, which the *Framework* describes as "reading situations," are included in the assessment. The situations are as follows:

1. Reading for literary experience, which includes the reading of novels, short stories, poems, plays, and essays.

2. Reading to be informed, which includes the reading of magazine and newspaper articles, textbooks, encyclopedias, special interest books, and catalogues.

3. Reading to perform a task, which includes the reading of bus and train schedules, directions for games, recipes, consumer warranties, and office memos.

Questions pertaining to reading for literacy experience could include "What is this story about?" and "How did Nancy change from the beginning to the end of the story?" Questions pertaining to reading for information could include "What caused the oil to spill in the sea?" and "What current event does this event remind you of?" Questions pertaining to reading to perform a task could include "What is this schedule supposed to tell you about?" and "Why do you need this information?"

The *Framework* discusses how the three situations are to be used as the basis for the development of the reporting scales, then makes recommendations for the proportion of items to be allocated to each situation.

Cognitive Aspects of Reading

To determine how well readers employ a range of cognitive abilities within each of the situations, this section of the *Framework* identifies four reading stances: forming an initial understanding, developing an interpretation, personal reflection and response, and demonstrating a critical stance. These four stances represent two cognitive aspects of reading: constructing the meaning of a text and elaborating and responding critically to it.

Constructing meaning implies understanding, in a general manner, what is read. This concept is based on the recognition that reading is a process that requires a reader to construct an understanding of the meaning of a text. Constructing meaning includes at least two of the aspects that have been identified: forming an initial understanding and developing an interpretation. The *Framework* advises that "while these abilities are related, it is possible to develop tasks for the assessment that focus on one or the other." So, a question assessing forming an initial understanding while reading to perform a task might be "What time does the bus leave for the courthouse?" A question assessing developing interpretation might be "What is the best route to take to get to the train station?"

The other aspect of reading, elaborating meaning and responding critically, requires readers to shift, consciously or unconsciously, to analytical reading. Analytical reading involves applying and judging the information or ideas from the text. So as to evaluate this type of reading, the *Framework* describes two broad categories of tasks, those that require personal reflection and response, and those that call for demonstrating a critical stance. A question assessing personal reflection and response in reading to perform a task might be, "To get to the courthouse by bus, what additional information do you need?" A question calling for demonstrating a critical stance might be "Why don't they include all the stops on the schedule?"

In identifying these aspects of reading--forming an initial understanding, developing an interpretation, personal reflection and response, and demonstrating a critical stance--the *Framework* emphasizes that they are not to be conceived of as a sequence or hierarchy. For example, a student might respond to a section of a text critically without developing an overall understanding. Further, while the stances are related and somewhat interdependent, some reading situations do not require students to engage in each one of the stances.

Constructing the Assessment

This section of the *Framework* discusses key aspects in the construction of the Assessment.

Designing items. The *Framework* calls for the use of a combination of open-ended and multiple-choice items and proposes that "the type of item used will be determined by the task and a commitment to increasing the use of open-ended items on this assessment." The rationale for open-ended items includes the need for having a means of looking at how readers integrate the reading of a passage with their own background knowledge and how they reorganize ideas and analyze and critically consider the text.

The section also explains that each of the open-ended items will be scored using primary trait scoring, and that scoring rubrics will be developed for each question. It gives directions for developing questions, both open-ended and multiple-choice, that will aid a student in building an understanding of, or examining the meaning of a text. Some of the questions have readers integrate information across passages. Item difficulty is determined by the difficulty of the passages and by the amount of knowledge the student must bring to the task to respond to the items.

Selecting passages. The *Framework's* discussion of passage selection implies a major departure from previous assessments. Passages were not to be written solely to assess a particular skill, rather, the *Framework* calls for the use of "authentic" texts, like those "found and used by readers in real, everyday reading . . . . Whole stores, articles, or sections of textbooks will be used, rather than excerpts."

The section advises that extended passages selected for inclusion be examined for coherence and orderly structure and with enough content so that they can be the basis for items that can lead to meaningful student performance. It further advises that teacher evaluation rather than conventional readability formulas be emphasized in establishing the difficulty level of passages, concluding that "the difficulty of text can be judged by the length of the text, the complexity of its arguments, the abstractness of its concepts, unusual point of view, and shifting time frames."

Special Studies

The rationale for special studies conducted with a smaller sample of the students is discussed in several sections of the *Framework*. The section of the *Framework* labeled "Special Studies" describes the Integrated Performance Record that includes two parts: the Oral Reading and Response Study and a Reading Portfolio Components Study. It also describes a third study of students' use of metacognitive strategies.

The Oral Reading and Response Study. In taped interviews, students are asked to read aloud and respond to items about a passage they have already read and responded to as part of the regular assessment. Their oral reading fluency will be analyzed by looking for evidence of their use of "phonics, sight vocabulary, semantics, and syntax." This information will be related to written and oral responses to questions about the passage.

The Reading Portfolio Components Study. The taped interviews will also be used to gather information about classroom reading instruction. For these portfolio-type activities, the students talk about both their independent and classroom reading assignments. In addition, they are asked to bring samples of their written work to the interview.

The Metacognitive Study. The *Framework* also set forth plans for a pilot study of the metacognitive strategies students use to monitor their reading comprehension. This study involves interviewing fourth-, eighth-, and twelfth-grade students to investigate the strategies they employ as they read.

BEST COPY AVAILABLE

As outlined in the introduction of this report, our principal charge from the National Academy was to investigate three issues:

1. Was the process used to develop the *Framework* adequate?
2. Does the *Framework* represent the consensus of the field of reading?
3. Does the assessment exemplify the ideas of the *Framework*?

What we did to investigate these issues, what we found, and our recommendations appear in the next section.

Issue 1: Was the Process Used to Develop the Framework Adequate?

To investigate the adequacy of the process used to develop the *Framework*, we conducted 50 personal and telephone interviews with representatives of a number of groups. These groups included the membership of the Planning Committee, a group of reading educators and administrators not directly associated with the development of the assessment, and some members of the National Academy of Education's Panel on the Evaluation of the NAEP Trial State Assessment Project.

The interviewers either taperecorded the conversations or took notes during the interview. A set of questions was developed to guide the interviews, but these were used flexibly as people responded to the topic (see Table 1). The information we gathered from these interviews revealed a number of strengths and some concerns about the development process.

[Insert Table 1 about here.]

Strengths of the Development Process

A major goal of the consensus process was to induce the interest and cooperation of key figures in reading and major professional education organizations. Given the time constraints, the project staff and Steering and Planning Committees did an excellent job of inviting broad-based participation, communicating the results of committee deliberations, and working toward a framework that would deserve broad support.

The *Consensus Report* best captures the techniques the project staff used to work under these time constraints:

> Many trips, meetings, and conference calls of committees were held on weekends to formulate and go over recommendations; we moved the entire planning committee to Palo Alto for a week just after new year's to inform the item and test specifications; we sent virtually nothing during the course of the project by regular mail; we conceived committees as rolling memberships from which we could draw at a given meeting, since there was no way to schedule a set of meetings when one set group of members would all be available. (pp. 16-17)

The efforts of the project coordinator were especially important to the success of the project. In addition to working on all aspects of the project with a great deal of energy and much determination, she made good use of her connections with professional reading organizations, the research community, and state reading coordinators. These connections made it possible for her to communicate directly with a number of people and groups to both give and get information about the developing framework and assessment. In the *Consensus Report*, the success of the *Framework* is attributed to "advisors and

planners who accepted tremendous awkwardness and inconvenience in order to do the work. Their willingness to work under these conditions made it possible" (p. 17).

The active interest and ready cooperation of the CCSSO staff member attached to the project should also be noted. He devoted a great deal of time as well as the resources of the organization to this effort.

The *Consensus Report* gives special praise to the efforts of the specification writers, and points to the importance of this type of coordination to any future efforts.

> Throughout the efforts and specifications team placed the continued objections paramount and dedicated themselves to faithful follow-through of their objectives. Never was it suggested that the concept of the assessment should be redirected or compromised to make it more convenient to assess or more in line with the traditional assessment practice. (p. 19)

Members of the Planning Committee that we interviewed also made positive comments about the development process. The following statements from three different members illustrate this point.

- I think that there was every effort made on the part of the measurement community to listen to the people in reading and I think if we had given better constructs they would have done things to support that. I feel like they have responded always with a great deal of respect to the people in the literacy community and been very responsive.

- There was a good balance between content people and measurement people. The content people didn't understand the measurement issues so it was important to have measurement people as well.

- The intention was good because they brought together lots of good people into the process. There was opportunity to react from a global perspective. The consensual process seemed to work. Given the time constraints things worked well.

Concerns about the Development Process

We have grouped our concerns about the development process according to seven topics: time, involvement of major professional organizations, state and local reactions, coordination of the development process, open-ended responses, and the quality of the public documents.

Time. A major concern is the small amount of time allotted for getting the work done and the gathering of consensus. Work on the *Framework* began in mid-October. The first draft was completed by January 30, at which time the work on the specifications document was well underway. The time allocated for the planning process, the development of the *Framework*, and consensus building was simply too short. Nearly every member of the project staff and the Planning Committee that we spoke to complained that there was too little time to do the work. These complaints were not about long days or interrupted schedules. Rather, the complaints were from professionals who were worried that some important issues had not been resolved, that some framework elements were internally inconsistent, and that incomplete specifications could lead to misleading assessment results. The *Consensus Report* captures the spirit of the problem:

> The project plan, membership of planning and steering committees, and background materials had to be completed in the first month, and virtually all project activities had to be scheduled in that first month.
>
> Obviously, this schedule resulted in casualties. Very little time was available for thoughtful recruiting of advisors, although appropriate experts representing the best expertise made themselves available. Background statements could not be circulated to experts for comment. No time was available for reflective informing of planning committee members or consultants. Drafts were circulated for a matter of days, rather than weeks or months. Materials were distributed to committees at meetings, sometimes, rather than ahead of time. Specifications were developed in parallel with the objectives. Reports had to capture the essence of a recommendation, rather than representing a careful, compelling statement of the position. (p. 15)

The short amount of time allocated for the consensus process was especially troublesome, given the known divisions within the reading community. The limited time given to each stage of the process, but especially for the creation of the *Framework*, meant that large numbers of people within the reading community had incomplete knowledge of and little involvement with the process. The *Consensus Report* states this most clearly:

> The biggest casualty was our inability to work effectively with the field. No time was available for reasoned circulation of materials and solicitation of response, and virtually no organized or formal response could be sought from organizations with a stake in the recommendations or with advice to give. (p. 15)

One of the members of the Planning Committee we interviewed indicated that the time frame may have been the reason why people from far-ranging perspectives were not included in the consensus process:

> I think there was minimal attention given to multicultural issues. Not many scholars from that perspective were considered in drafting the framework.

Another member remarked,

> I'm not sure how satisfied I am with the process used to develop the framework. I am satisfied with my participation but having so many constituencies was cumbersome. It was probably necessary to have members from all the groups who might have a vested interest in the NAEP Reading.

On the other hand, in acknowledging the importance of NAGB consistently backing the project to "range fully and think openly and widely about assessment" and not be "constrained or compromised by past practice or current resource constraints," the *Consensus Report* concurs that the project was "consistently able to demonstrate to the field that the planning effort was sincere and uncompromised" (p. 14). We remark that this statement seems optimistic, particularly in light of the decision of the International Reading Association (IRA) to withdraw from any involvement with the assessment.

Involvement of major professional organizations. The pressures from legislative and other public groups for accountability can place professional organizations in a difficult position. As the *Consensus Report* states "on the one hand, they [professional reading organizations] should represent their members' concern about proliferating and misused assessment. On the other, opposing politically popular assessment programs can make them appear to be avoiding accountability" (p. 11).

The decision of IRA to disassociate itself from the process was particularly troublesome. To illustrate this point, some background is necessary. More than a year before the Reading Consensus Project began, IRA adopted a resolution in opposition to the "proliferation of inappropriate assessments." IRA's Board of Directors interpreted this resolution as precluding both the organization's support for the assessment and its official involvement in planning the assessment. This position meant that the Reading Consensus Project could neither use official IRA resources nor seek IRA positions on specific issues. As the *Consensus Report* notes:

> It is not clear in hindsight that IRA's interpretation of its position was necessary. Its opposition to "inappropriate assessments" was not necessarily in conflict with the goals and principles of the 1992 Reading Assessment planning effort. Indeed, IRA, by participating, could have helped insure that the plans resulted in an appropriate assessment. The project and the organization were not that far apart from one another in their goals and values. (p. 12)

The question remains whether IRA would have been involved under any circumstances, or whether it chose not to be involved because of the limited time available to the Reading Consensus Project. In any case, dialogue that might have established a shared commitment with IRA did not occur. It is of interest to note that in May 1991, the IRA Delegates Assembly approved a set of four resolutions on literacy assessment. These new resolutions support IRA's involvement in new forms of assessment as long as these assessments are treated as "experimental, purposeful, flexible and respectful of differences among students" (International Reading Association).

Another major organization, the National Council of Teachers of English (NCTE), took no official position opposing the assessment. However, at its annual conference that occurred during the project's planning period, NCTE's Commission on Reading did take a position questioning the assessment. It was not clear how far this action placed NCTE in opposition to the project, and the position itself was later changed to one that was far less antagonistic to the process and, according to the project coordinator, somewhat supportive of the assessment.

State and local reactions. Another factor affecting support for the assessment was the wariness of several directors of some state and local testing programs. Several states had gone through the arduous process of developing new directions for their reading curriculum and assessment programs. Their concern was that the state reporting aspect of the assessment, with its high visibility and anticipated impact, would challenge and threaten the directions they had set and the progress they had made in those directions. The comments of a Planning Committee member who works for a state board of education are particularly relevant: "I would like to have seen more [state] administrators involved in the process because they are the ones who will be making policy decisions."

Still another factor that affected support from the states was that NAGB had made long-term changes to the NAEP program that included considering that the prohibition against using NAEP data for school and district comparisons be dropped. According to the *Consensus Report*:

> The prospect of these changes alarmed some state and local test directors . . . , requiring effort by NAGB and the project to assure those with a stake that such changes would not be made without due consideration and an opportunity by those who would be affected to comment. For the project, it meant some difficulty keeping key players "at the table," so the consensus planning effort could be completed. (p. 13)

In constructing the *Framework* for the 1990 NAEP in mathematics, the content of state and local curricula and assessment policies and objectives was systematically analyzed. This was not the case for the reading assessment. As is explained in the *Consensus Report*:

> No such review was possible in reading. Although materials on state reading curriculum and assessment were available to committee members, there was not sufficient time to review and digest these materials. Objectives for the national assessment were formulated without any real sense of the breadth and variety of state and local curricular emphases. (p. 16)

Coordination of the development process. Our next concern has to do with the coordination of the various aspects of the total development process. The development process included activities such as goal setting, planning, framework development, item-specification writing, item writing, development of scoring rubrics, field testing, scoring, and reporting. In some cases, there was close coordination of the activities, for example, between the writing of the framework and the specifications documents. But even in this case, because of the severe time constraints required to meet the deadlines for the completion of these documents, each had to be completed in the same month--thereby precluding an orderly coordination between the objectives and the assessment specifications.

From the beginning, the requirement was that the *Framework* and *Specifications* be done at the same time. The *Consensus Report* supports this feature of the charge:

> This feature had many advantages. It resulted in thinking about the assessment objectives which was much more specific and concrete than it would have been without this task. Directions and implications to the test developer are much more clear and unambiguous. Planing for implementing the recommendations began much earlier, allowing for more orderly handling of logistical, funding, and procurement issues involved in carrying through on the recommendations. Finally, the methodology of the assessment could be advanced more effectively, because assessment methods had to be thought through for the recommendations at an early stage. (p. 18)

But the *Consensus Report* suggests that a longer time period would have permitted a more appropriate phasing of the *Framework* and the assessment, "it was simply not reasonable to complete the objectives and specifications in the same month" (p. 18). The report proposes that in the future, a two- or three-month lag between the deadlines for framework and specifications documents would allow more orderly attention to each task.

Open-ended responses. The concern here is with the evaluation of the open-ended responses. As is obvious, the scoring of open-ended responses is very different from the scoring of multiple-choice items drawn from an item pool. Because content and pedagogical knowledge are critical to the item creation, scoring, and interpretation of open-ended responses, the research and teaching communities should be involved in this process.

The quality of the public documents. Finally, there was a concern that the *Framework* shows evidence of being written in haste by a committee. Given the time constraints for its completion, and the number of people who contributed to it, this should come as no surprise. For example, there are some inconsistencies within the *Framework* and between it and the other documents. The explanation of the concepts associated with constructing, extending, and elaborating meaning needs some clarification. The organization of its various sections would be improved by more specific headings and subheadings. This is no list of references to inform the reader about the origin of the ideas it proposes as the basis for the assessment. Having said all of this, we want to acknowledge the *Framework's* assets. As one Planning

Committee member remarked, "The framework is comprehensive, has a lot of positive changes, and is a miraculous document in light of the time allotted."

NAGB responded to these concerns by revising the *Framework* to provide a clearer rationale for the 1992 NAEP in Reading.

Recommendations about the Process Used to Develop the 1992 NAEP in Reading

Given the time constraints under which the project staff and the Steering and Planning Committee operated, we were impressed with the efforts made to involve reading educators and professional literacy organizations in the process of developing the 1992 NAEP in Reading. Participation was broad-based and opportunities for communication among interested parties was ongoing. Based on the information we gathered through a series of personal and telephone interviews with members of representative groups, we offer five recommendations:

1. There should be a *center* of responsibility to oversee all aspects involved in developing, administering, scoring, and reporting a NAEP in Reading.

2. There should be closer coordination among the institutions and groups involved in the various aspects of developing and implementing the NAEP in Reading.

3. The involvement of people from the field of reading should extend beyond the planning stages. They should participate, for example, in making decisions about scoring and reporting.

4. More time must be allocated for the planning process. Without time to consider the diverse viewpoints within the field of reading, it is not possible to build a wide consensus.

5. The documents produced by the committees for public dissemination should be clarified and made more consistent.

Issue 2: Does the Framework Represent the Consensus of the Field of Reading?

To investigate the adequacy of the *Framework*, we each read, studied, and outlined the *Framework*, and then discussed its content in several of our own meetings. We also presented its content to two larger groups--participants in two colloquia and members of a panel we convened. We also incorporated the content in a survey questionnaire. A detailed description of each of these activities follows.

We sent announcements of our two colloquia to approximately 200 people, primarily staff members of the University of Illinois at Urbana-Champaign and of some neighboring institutions. These colloquia were held February 2 and 15, 1991, at the Center for the Study of Reading. Approximately 50 people attended each session; many of the same people came to both. Each session was audiotaped, and these tapes were later transcribed. Information presented in this section of our report was drawn from the first colloquium. The activities of the second colloquium will be discussed in a later section.

In the first colloquium, we presented background information about the process used in developing the *Framework* and an outline of its essential features, especially the categories implied by the three reading situations and the four stances of reading it identifies.

On April 1 and 2 at the Center, we convened a panel comprised of three professors of education, one member of a state board of education, one public school administrator, one leader from IRA, one leader

BEST COPY AVAILABLE

from NCTE, and one educational consultant. We deliberately excluded from this panel people who had been involved in the development of the *Framework* or the assessment.

In advance of the meeting, the panel had been sent copies of the *Framework*, the survey questionnaire, and our proposal to the National Academy. The first day of the meeting was devoted to an extensive discussion of the *Framework*. The second day was devoted to consideration of the match between the *Framework* and the assessment. This day's activities will be discussed later. Both days' proceedings were taperecorded and later transcribed.

Is the *Framework* a document of consensus for the field of reading? The answer to this question is not simple. To attempt to answer it, we will first discuss the strengths of and then concerns about the *Framework*, as identified by the colloquium participants and members of the panel. Then, we review the responses to the survey questionnaire, a document based on the *Framework*, which was sent to a national sample of 700 educators.

Strengths of the Framework

Most of the participants in the colloquia and the panel agreed that the *Framework* has a number of important strengths. There was almost unanimous approval for a number of the features of the assessment. These include that the 1992 NAEP in Reading is an approach that:

Aligns with the process of reading. Many of the participants seemed surprised and pleased with the general approach taken in the *Framework*. They approved of its claim that "the 1992 design builds on recent studies to view reading as a constructive, dynamic process, not just the assembly of a set of subskills." Among our respondents, there was broad agreement that the assessment is strongly aligned with what is known about the process of reading. One researcher said, "I'm pleased to see greater attention [compared to traditional tests] being paid to what is being tested." Another participant commented, "I think the *Framework* itself much more closely reflects what both research and practice are doing right now, in describing reading as an instructive, interactive, complex process. This assessment represents a tremendous step forward. It's a lot stronger basis for assessment than we have had in the past."

Use of open-ended questions. The *Framework* calls for approximately 40% of student time to be spent responding to open-ended questions. This change was looked upon very favorably by our colloquia and panel participants. A second point of broad agreement was that the 1992 design's emphasis on open-ended responses was a major advance, and most people applauded this effort to seek more elaborated responses from students. One panelist pointed out that this approach "makes reading visible" for the first time on such a large-scale assessment. There was also, however, a recognition that scoring open-ended responses presents a fairly formidable challenge in large-scale testing.

Uses authentic texts. The *Framework* calls for the assessment to use naturally occurring, whole, "authentic" passages rather than isolated words, single sentences, or passages written especially for testing purposes. Authentic passages are longer and are generally more challenging and more interesting than the more typical assessment passages. Most important, they more closely approximate the kinds of reading students engage in at home and in school and thus represent a more "ecological evaluation" than the specially written test passages (Lucas, 1988a, 1988b). Virtually all of the participants agreed that the use of authentic texts marks another advance in the assessment.

Allows student choice in texts. The *Framework* calls for twelfth-grade students taking the assessment to choose one story from a booklet containing several stories. It was a surprise to many participants that such an approach would be tried in a large-scale assessment. The idea that students have choices in what they read for the assessment was also considered to be an advance.

Concerns about the Framework

There were concerns about the *Framework* expressed by the participants in the colloquium and members of the panel. We discuss these in this section.

Reading situations. The assessment accords with recent reading research in focusing on the situation in which reading occurs. The situation is important in shaping the reading process and the construction of meaning from a text. An obvious concern is that the passages and items on the assessment do not represent authentic reading situations. For example, a student taking the assessment is not, in fact, "reading for a literary experience," "reading to be informed," or "reading to perform a task," but rather he or she is "reading to take a test." Thus, at best what the assessment can do is approximate the other reading situations, through the type of passage, the questions, and the format.

Another problem with the reading situation focus is that it implies that there is a one-to-one correspondence between text genres and reading situations. Thus, within the assessment, students read fictional stories only "for literary experience" not "to be informed." This one-to-one correspondence contradicts the idea of "reading situation" as the term is usually employed in reading research, which conceives of the situation as defined by the reader's purpose, the social context, the task, and only in part by the text. One person we interviewed said, "I think they have confounded purposes for reading with genre, and that is confusing. I would rather see them stick with genre."

The implied one-to-one correspondence can have negative pedagogical consequences, especially if it suggests that this is a test worth teaching to. A teacher might infer that students should not be encouraged to learn from, or be informed by a literary text. This is exactly the opposite of recent recommendations about using reading in the content areas. For example, Butzow and Butzow (1989) show how children's literature can be used in the teaching of a variety of science topics. As one colloquium participant noted, "I don't understand how reading literature to explore the human condition is not reading to be informed."

Another question is whether the reading situations are as "authentic," as the *Framework* claims. Testing requires students to read in isolation and give relatively short responses, with no opportunity to revise those responses. While this is certainly one kind of reading, it is by no means representative of the many kinds of reading situations in which students engage or that reading researchers recommend.

Operationalization of stances. An important experiment in the assessment is the attempt to assess the ability of readers to adopt different stances with respect to a given passage. This view of reading originates in literary theory, especially reader-response theory (Fish, 1980; Tompkins, 1980). In contrast to the picture of the reader as one whose job is to glean information uncritically from the text (a special case of what Rosenblatt, 1978, calls *efferent* reading), we now have a picture in which the reader assumes not only different purposes, but different relationships to the text and the author. These relationships have been identified as stances.

Stances can be conceived of in various ways. One is to see a stance as a personal relationship to a task environment. Thus, Hartman (1991) found that readers adopted different stances as they read a set of texts based on their own interests and purposes. Other models depict stance as a fluid relationship that emerges through readers' ongoing construction of meaning. In her analyses of readers' think-aloud reports, Langer (1990) found that as students developed their meanings across time, the ways in which they related to the text (their stances) changed, with each stance adding a somewhat different dimension to their understanding of the entire piece. Yet another model using something similar to stances is the interpretive community idea suggested by Fish (1980). Here, a reader's relationship to the text is not determined by the text, but it is not entirely personal either. Instead, it emerges from the reader's

participation in an interpretive community. Thus, a classroom of students studying poetry may be so primed to interpret texts poetically that they read a list of names as a poem.

The descriptions are but a few of the many theories that have been proposed for describing the reader's relationship to a text. The assessment uses stances to develop questions, and in that sense is responsive to an increasing interest within the reading community to considering how the reader interacts with a text. But to assess responses, it uses a model somewhat different from any of those mentioned above. Instead of seeing stance as a response that varies widely among individuals, including highly-skilled readers, or even during the reading of a passage, the assessment sees stance as imposed by the task. Thus, the combination of a passage and a question causes the reader to adopt a particular relationship to the passage. This apparently subtle shift in definition to meet assessment needs leads to a very different model for stances, one that was problematic for the participants in our colloquium, the members of our panel, and us. Fortunately, as we understand it, the assessment will not report the stance data.

One researcher at the colloquium put it this way:

> I think a number of these are not bad questions. They are the sorts of questions I would hope children would be able to answer. My particular complaint is . . . that the relation of the questions to this *Framework* is extremely murky and insofar as they are supposed to be an instantiation of this *Framework*, I think in general they failed, I think there are very few where there is a clear-cut relationship.

This statement reflects one of the guidelines given to the Planning Committee, that is, that the *Framework* should be a document that "focuses on outcomes or is performance oriented, rather than reflecting an instructional or theoretical approach."

We question, however, whether a solely performance-oriented assessment is, in principle, possible. Any reading assessment rests upon a number of assumptions about what reading is, how it relates to other aspects of learning, how readers interact with texts, how they respond to what they read, and what role instruction can play in learning to read. For example, existing reading assessments employ everything from lists of nonsense syllables (as in a test of word recognition) to full-length novels (as in a portfolio assessment). Similarly, assumptions about reading shape what counts as an appropriate measure of reading--the choices are almost endless, everything from the tracking of eye movements to a call for artistic responses such as drama or painting. Theoretical justifications can be made for selecting from these different choices. A particular theory or its relationship to the choice of text types can be contested or embraced, but we know of no way to make such choices independent of a theory base, even if that theory is not well-articulated.

We believe that the *Framework* does represent a set of values and beliefs about reading that are derived from extensive research on reading and reading assessment, and that, in fact, the ideas about reading that are expressed in the *Framework* are based in theory. (See Anderson et al., 1985; Langer, 1989, 1990).

The Survey Questionnaire

The content of the *Framework* was used to prepare a survey questionnaire that would bring us reactions from a larger and more varied sample of educators. We discuss these reactions in this section. Because both the strengths of and concerns about the *Framework* are reported in the survey data, we present them together in this section.

Project staff developed the questionnaire, working closely with the University of Illinois Survey Research Laboratory. In April 1991, we mailed approximately 700 questionnaires and received 308 responses. (See Appendices A, B and C for the cover letter, the overview of NAEP, and the survey questionnaire that were used.)

We sent the questionnaire to (a) a random sample of 250 participants listed in the program for the annual meeting of the National Reading Conference; (b) a random sample of 250 Presidents of IRA Councils, drawn from the 1990-91 Desktop Reference to the IRA; (c) all 50 chief state school officers; (d) all 50 state directors of Chapter 1 reading programs; (e) all 50 state reading specialists; (f) everyone who had responded in writing to earlier drafts of the *Framework*; and (g) leaders in IRA and NCTE. Due to time constraints, we did not do a follow-up mailing.

In August we sent an additional 1,000 questionnaires to classroom teachers who were members of IRA. We received approximately 300 responses. Details of this second survey are reported in Commeyras, Osborn, and Bruce (in press).

The analysis to follow focuses on the initial survey sent to 700 educators. More than half (58%) of the respondents to the first survey indicated that they were employed as teachers at the elementary, secondary, college, or university level. About half of this group held college or university positions.

Some 37% of the respondents held administrative positions at the school, district, or state level. Almost 70% of the respondents had more than 15 years of experience in education. More than 85% of the respondents felt they were somewhat or very familiar with NAEP. (See Section K in Appendix C for more detailed information regarding respondents.)

Overall the responses represent a great deal of support and agreement with the contents of the *Framework*. The following discussion provides some specific information about the survey results. We organize this discussion according to the headings used in the questionnaire.

Characteristics of good readers. An overwhelming majority (93% or more) of the respondents agreed with the characteristics of good readers presented in the *Framework* (see Section A in Appendix C). There was complete agreement expressed for the following two characteristics:

> Good readers read with enough fluency so that they can focus on the meaning of what they read, rather than devoting a lot of attention to puzzling out the words.
>
> Good readers use what they already know to understand the text they are reading.

A negligible percentage (less than 5%) of the respondents disagreed with the other characteristics listed.

Views of reading. The *Framework* contains a number of statements that represent definitions of the view of reading that were to be assessed in 1992. In our survey, we asked respondents to indicate the extent to which they found four of these definitions acceptable (see Section B in Appendix C). More than 90% indicated that they found the four definitions of reading either acceptable or very acceptable. The following definition of reading received the strongest support (85% of respondents found it very acceptable).

> Reading is a complex process that involves an interaction among the reader, the text, and the context, or situation.

BEST COPY AVAILABLE

Whereas only 68% of respondents selected the following definition as "very acceptable."

> Proficient reading contributes to a sense of personal satisfaction.

Reading situations. Respondents agreed that it was important to assess students' reading ability in the three situations identified in the *Framework* (see Section C in Appendix C). More than 80% of the respondents thought it was either very important or absolutely essential to assess reading in these three situations. It is interesting to note that more than 76% thought it was absolutely essential to assess students' ability to read to be informed and to perform a task, while only 43% thought it was absolutely essential to assess reading for literary experience.

The survey also sought reactions to the proportion of items allocated at each grade level to each type of reading situation. More than half (59%) of the respondents indicated that they disagreed with the allocation of items. There was a great deal of variability in the percentages of items for each grade level and reading situation suggested by the respondents.

Cognitive aspects of reading. The *Framework* identifies four stances (forming an initial understanding; developing an interpretation; personal reflection and response; and demonstrating a critical stance) that are representative of some cognitive aspects of reading. These stances were to be assessed in each of the three reading situations. Most of the survey respondents (more than 88%) thought it was either absolutely essential or very important to assess these (see Section D in Appendix C). Sixty-nine percent of the respondents said that there were no other aspects of reading that needed to be represented in the assessment.

Open-ended items. The respondents were extremely supportive of the inclusion of open-ended items in the assessment (see Section E in Appendix C). Only 3% objected to having open-ended items. Approximately 87% of respondents agreed with the rationale for including open-ended items. Approximately 12% of the respondents thought that 40% was too much to allocate to open-ended items on the assessment. Over 62% were comfortable with this percentage.

Passage selection. Considerable support (66%) was shown for the decision to only use authentic passages. Interestingly, 30% of the respondents favored a combination of authentic passages and passages written to test specific skills (see Section F in Appendix C).

Teaching to the test. We asked respondents to indicate whether NAEP should attempt to develop an assessment that could serve as a useful guide to instruction (see Section G in Appendix C). Only some 26% of respondents opposed this idea.

Special studies. We sought reactions to the special studies that were to be conducted with a subsample of students. Respondents were most favorably disposed toward the portfolio assessment and metacognitive study (see Section H in Appendix C). More than 50% believed these two studies were important. They were less enthusiastic about the oral reading study. Only 15% thought such a study was needed to a very great extent.

Goals of the 1992 NAEP in Reading. Survey respondents were asked to judge the extent to which the *Framework* met the guidelines that had been set out by the Steering Committee. The majority of respondents indicated that the *Framework* seemed to meet the five guidelines (see Section I in Appendix C).

State-by-state reporting. The respondents had very different reactions to the move toward state-by-state reporting. Approximately 61% were either strongly or moderately in favor of this reporting. Another 36% were either somewhat or strongly opposed (see Section J in Appendix C).

Issue 3: Does the Assessment Exemplify the Ideas of the Framework?

To examine the passages and items of the assessment and to determine the degree to which they exemplify the *Framework* and the *Specifications*, we engaged in several efforts. First, we analyzed and categorized the passages and items that ETS had prepared for the 1991 field test. Next, to find out how some experts in the field (who had not been associated with the development of the *Framework*) would compare the assessment to the *Framework* and the *Specifications*, we turned to the participants in the second colloquium held at the Center for the Study of Reading and to the members of the panel we convened. The colloquium participants examined and categorized the sample passages and items that had been included in the *Framework*, whereas the members of the panel did the same thing, but with some of the passages and items from the ETS field test materials. We discuss each of these efforts in turn.

The In-House Analysis

To compare the content of the *Framework* and the *Specifications* with the content of the passages and items, we analyzed a sample of 10 blocks of passages and items from the three grade levels (fourth, eighth, and twelfth). A block represents what a student reads and responds to in one test session. It should be noted that these released items are similar to, but not identical to the actual test items.

For our analysis, we selected blocks of passages and items from those that had been developed by ETS for use in the 1991 field test. These were chosen to be representative of the grade levels and the three reading situations. Our sample comprised the following:

1. Grade 4: 2 literary and 2 informational blocks
2. Grades 4 & 8: 1 informational block
3. Grade 8: 1 literary and 1 informational block
4. Grade 12: 1 literary, 1 informational, and 1 document block

The analysis involved checking the item distribution for each block with the exercise descriptions detailed in the *Specifications*. In addition, we categorized the items in each block according to the four stances set forth in the *Framework*. Finally, we compared our categorization of the items to that of ETS to ascertain the extent to which they agreed. In the materials ETS developed, a list at the end of each block matched each item with one of the four stances. How well the items represented the reading stances was a major concern. Therefore, we undertook the task of backcoding the 10 blocks of test items to gain insight into the extent to which specific test items corresponded to the four stances specified in the *Framework*. Table 2 shows the results of this backcoding.

[Insert Table 2 about here.]

As the table shows, the rater had considerable success backcoding items for three of the four stances. This consistency emerged as the rater became accustomed to the match between descriptions of initial understanding, developing interpretation, and personal response with the corresponding items. The most serious difficulty arose with items categorized as critical stance. According to the *Framework*, demonstrating a critical stance requires a reader to "stand apart from the text and consider it objectively," and then engage in "critical evaluation, comparing and contrasting, application to practical task, and understanding the impact of such text features as irony, humor, and organization." Our analysis showed, however, that most of the items intended to assess students' ability to take a critical stance did not (from the perspectives of our rater) require the reader to consider a text objectively.

In fact, many items designated by ETS as critical stance seemed to fit more appropriately one of the three other stances. For example, a twelfth-grade critical stance item asks: "How does the author build tension or excitement in the story?" This item does not necessarily compel the student to stand apart from the text and consider how the author builds tension. Instead, the student could point to one event in the story that built tension and therefore not take a critical stance. The following rephrasing of this item would seem to fit more closely the *Framework* description of a critical stance item: "Defend your view of the extent to which the author effectively builds tension or excitement in the story."

We analyzed the sample of item blocks for accuracy, that is, to determine if ETS's accounting of the items for each of the stances was accurate. We also looked at the distribution of open ended and multiple-choice items to check if ETS followed the guidelines for multiple-choice and open-ended responses.

We found the ETS accounting of the number of items per stance to be accurate; discrepancies were rare and were usually typographical errors. We did encounter a formidable problem when we tried to determine whether ETS had provided the appropriate proportion of open-ended and multiple-choice items. The *Specifications* deal with the distribution of items according to the time students spend doing them. According to the *Specifications*, students should spend 40% of their test time on open-ended items and 60% of their time on multiple-choice items. We could find no estimation of how much time ETS believed it would take to complete the open-ended and multiple-choice items. In counting the *number* of open-ended and multiple-choice in each block, however, we found that our sample of 10 blocks contained approximately 48% multiple-choice and 62% open-ended items. Because the *Specifications* refer only to *time* to be spent, and we could only count items, it was difficult to determine if ETS has provided the correct proportion of open-ended and multiple-choice items for the assessment. It seems strange that 62% of the items in our sample were open-ended, and yet according to the *Specifications*, students are only supposed to spend 40% of the test time completing them.

The Colloquium Analysis

Participants in the second colloquium were asked to categorize the items from the sample passages and items that accompanied the *Framework* according to the three reading situations and the four stances. They "took the test" with the block of items accompanying one passage, and also discussed the items that were used in the *Framework* as examples of each of the stances. What follows is a brief review of the assessment's strengths, as well as the concerns about it that emerged from the discussion.

Responses other than writing. The colloquium participants agreed that the assessment is unusual in its incorporation of writing as a response mode for reading, and therefore represents a significant advance in large-scale assessment. But several pointed out that other response modes should also be considered. For one participant, it seemed odd to have only multiple-choice, short-answer, and essentially literary forms of writing responses for items to assess "reading to perform a task." The participant argued that none of these responses correspond to the real-world way people read to perform a task.

Participants suggested other response modes be considered for future assessments. For reading to perform a task, these could include setting up an apparatus for a scientific experiment, using a computer to edit a document or analyze data, locating information in a library (or at least in an encyclopedia), or writing a resume from a fictional biography.

A second reason for including responses other than writing stemmed from concerns about confounding writing ability with reading ability. One participant commented that

> You have to be a good writer, know how to write a good paragraph, in order to demonstrate comprehension. So it's mixing the two. Second language kids are not going to be able to do this. Some second language kids can read quite well and comprehend but are not going to be able to write what they are expected to.

This view was shared by a researcher, who asserted, "I am very concerned about measuring reading through writing. I think that is going to be very misleading because anyone who has been having trouble with writing is going to score low on reading."

Item distribution. A great deal of discussion concerned the item distribution across reading situations and grade levels. The *Framework* calls for different percentages of items across text types for each grade level, as shown in Table 3.

[Insert Table 3 about here.]

Several members disagreed with this distribution of items. These views are best expressed in the remarks of one participant:

> The number of items you include should have nothing to do with how much reading is done [in school]. You just need enough items to get a valid reliable measure of that particular behavior. I don't understand this at all. It could be 13 items, 13 items, 13 items. However many it takes to get an index of how well you can perform a task.

The colloquium participants realized that the Planning Committee for the *Framework* wanted not only that the individual passages be authentic texts but also that the relative numbers of passages of each type be authentic representations of what students do in schools. That is, if a type of passage represented 10% of what students read, it should comprise 10% of the test items. Several participants argued, however, that these laudable concerns for authenticity can lead to other problems. By not having equal numbers of items in each cell, the reported results will differ in their degree of uncertainty. For example, the field can have less confidence in the accuracy of the twelfth-grade reading-to-perform-a-task results (20% of the items for that grade level) than in the fourth-grade reading-for-literary-experience results (55% of the items for that grade level). Thus, there is also increased uncertainty in the entire set of results for little, if any gain. One participant pointed out that it would be strange to argue that a category of performance is important enough to be tested and reported but not important enough to have reliable results.

Fourth-grade reading to perform a task. Another concern identified was about the Planning Committee's determination that fourth graders spend less than 20% of their reading time on reading to perform a task. The committee members reasoned that, with fewer than 20% of the fourth-grade items being in that cell, the confidence level would be too low. So they decided not to assess fourth graders' ability to read to perform a task. The difficulty with this decision is that the assessment becomes truly unauthentic, because everyone agrees that fourth graders do read to perform tasks. As one participant put it:

> A real concern is that reading to perform a task includes following written directions, which is what a great deal of school work is all about. So it should be a concern for a national test at fourth grade.

The colloquium participants suggested a conflict between concerns for authenticity, "teachableness," and sampling reliability. It seemed preferable to them to assess important reading abilities reliably, using a uniform distribution of items across grade levels and situations.

The Panel Analysis

Like the colloquium participants, members of the panel examined passages and items. In the case of the panel, however, the passages and blocks were from the materials developed by ETS for use in the 1991 field test. In conjunction with "taking the test," the panel analyzed and backcoded the items for three blocks.

The results of the panel's backcoding provided further insights about the use of the four stances of reading in categorizing the items. In many cases, the members of the panel thought an item represented a different stance than the one the item writers had assigned it (see Table 4).

[Insert Table 4 about here.]

Approximately half of the time the panel members agreed with the categorization assigned to items belonging to initial understanding, developing interpretation, and personal response. The percentage of agreement for items in the critical stance category was especially low. Note that these percentages are generally lower than those reported in our in-house analysis in Table 2. The results of our in-house analysis shows that there is some coherence, while the panel's results show that this coherence is not transparent and may only be realized after close study. The difficulties our panel members had backcoding the three blocks of test items they examined is troublesome. The following discussion looks at specific concerns raised about each category.

Initial understanding and developing interpretation. There was considerable agreement among panel members that the initial understanding and developing interpretation categories were mislabeled and thus confusing. One panel member explained her concern this way:

> I had a problem with the term initial understanding. I think of initial understanding as when you are trying on different schemas to see where the author is going but that's not what they mean in the framework so I think the word initial is misleading.

Another panel member suggested that the description of the initial understanding category would make more sense if it were called "global or overall understanding." We found that the panel members were apt to label items that had been categorized by ETS as initial understanding, as developing interpretation and vice versa. In sum, the panel thought that the category labels of initial understanding and developing interpretation did not fit their descriptions offered in the *Framework*.

P rsonal response. The panel was pleased to see the inclusion of items that called for a personal response, but they were concerned about the scoring procedures. They objected to scoring guides for personal response items that provided specific information to be included in "correct" answers. One panel member explained the difficulty this way:

> When you look at the scoring guide you find out that it isn't personal response at all. You can't call an item personal response when the examinee has to include two of the ideas specified in the scoring guide. That immediately takes it away from being a personal response. Personal response means you ought to be able to respond to the question in any way you choose. There's no way that anyone could score these things without an interview. I'm hard pressed to mark anything as personal response in this set.

This concern was discussed by other reading specialists that we interviewed. For example, one said,

> With the personal response items, one must accept any reasonable answer. If everyone gets these items right, what are we assessing? It will add points to each student's score. How meaningful is this data and how is it going to be reported?

Critical stance. An extended discussion occurred among our panel members about the items classified as demonstrating a critical stance. The basic issue was whether the items matched the description offered in the *Framework*. According to the *Framework*:

> Demonstrating a critical stance requires the reader to stand apart from the text and consider it objectively. It involves a range of tasks including such behaviors as critical evaluation, comparing and contrasting, application of practical tasks, and understanding the impact of such text features as irony, humor and organization.

The panel discussed this description in conjunction with a number of critical stance items. Many of these items were categorized by panelists as initial understanding, or developing interpretation. One panel member explained why an item did not seem to call for demonstrating a critical stance.

> I don't see that the reader is being asked to stand back and consider the quality and organization of the author's ideas and presentation of information. The reader is simply being asked to identify a statement that is valid given the textual information. There is no evaluating quality. The reader is just identifying what the author has done.

Yet another member of the panel offered a possible explanation for why the item was classified critical stance.

> I'd like to offer a possible defense for their categorization but I don't accept it. In saying that a statement is supported is to endorse the relationship between the evidence and the conclusion. To say there actually is support there rather than no support calls for some evaluation. But I think it is so obvious that I wouldn't go along with it.

One of the panel members, who is an expert in critical thinking, suggested that the following question types would prompt items that might better fit with the description of critical stance given in the *Framework*: Is the author's conclusion justified? What is being assumed? Are the author's assumptions valid? Is the author credible? and What is the author's perspective?

Our panelists' concerns about critical stance were echoed by two members of the Planning Committee. One commented,

> The personal reflection items are going to be very hard to score. When kids bring in some personal reflection sometimes and you are asking them to base it on something that happened in the text there is some real conflict there in scoring it.

And another remarked,

> I hope they will examine the relationship between open-ended and multiple choice items to examine the cultural implications of including open-ended items. It may be that in some cultures there is a propensity for verbosity while in others succinctness is valued.

Open-ended responses. The panel had concerns about the time allocated for open-ended responses. To reduce obtrusiveness to schools, the assessment is administered in 25- and 50-minute blocks.

One problem is associated with the booklet of stories. One panel member noted that a student could easily spend 10 minutes deciding which story to read, particularly if that student had spent time in a classroom environment that encouraged choice, and might then spend 20 minutes reading a story, more if she were a slow reader or a fast reader who wanted to re-read a passage or stop to think about what she was reading. That would leave less than 20 minutes for 12 open-ended questions, little more than 90 seconds a question.

Thus, the time for responding would be short, which is not what most teachers and researchers envision when they speak of encouraging open-ended responses for assessment purposes. Given the amount of time available for the assessment, our panel's strong consensus was that there should be fewer questions. The members preferred to see more detailed analyses of fewer, more elaborated responses, rather than cruder analyses of many short responses.

Other questions. In addition to discussing the match between the passages and the items and the *Framework*, the panel members raised a number of other questions, including the following:

1. Are the passages appropriate? Several people thought the choices available in the twelfth-grade stories was peculiar -- in the word of one panel member, "odd."

2. Why is student choice limited? Another panelist asked why student choice is limited to reading for literary experience and not extended to reading for information and reading to perform a task, areas in which prior knowledge and interest vary enormously -- and would presumably affect performance.

3. Why are generic questions used? Several members of the panel expressed dissatisfaction with the generic questions written to be used with all of the stories the students choose from the booklets. The point was made that these questions were often a poor match for the individual stories.

4. How will the written responses be scored? The advise was that the scoring categories should be derived from the data. For example, ETS should use situations and stances as hypotheses to be revised on the basis of student responses on the pilot test.

5. What about primary trait scoring? Several panel members felt that primary trait scoring of the open-ended responses may not be appropriate for a diverse student population, especially considering the emphasis on interpretative, personal, and critical responses.

6. Will the scoring procedures confound writing performance with reading? Almost all of the panel members were concerned about the effect of writing competence on the evaluation of reading. This was especially a concern when considering the challenges the assessment poses to minority students, most especially LEP students. The panel urged that a great deal of time after the field test be spent dealing with this problem.

7. Will the assessment be an advance in large-scale testing? The panel members' responses to the question ranged far. For example, one member described it as "a bold experiment," another as "more of the same," and still another as "a possible step backwards." One member's concern was that it "lagged behind classroom-based

research on portfolio and situated assessment." But another liked it because he considered it "a check on the national testing movement."

8. What are the limitations of large-scale assessments? Members of the panel urged the importance of recognizing from the outset some of the inherent limitations of large-scale assessments. They cautioned that the 1992 NAEP measures only some aspects of learning, does not represent integrated learning, and is segmented by subject area. The limited time allocated for students to take the assessment is, of course, constraining. Individual students receive only a small sample of passages, there is little or no chance for revision, and no chance for student-teacher or student-student collaboration. Finally, it was pointed out that even with the open-ended responses, the response formats are still restricted. The point was made about the danger of demanding from a large-scale assessment tasks for which it is inappropriate. This could lead to either unwarranted criticism of the assessment or, what's worse, attempts to apply it to tasks it was never intended to carry out.

Our findings from our investigations of the correspondence between the *Framework* and the assessment lead to three groups of recommendations. These appear below:

Recommendations about the Adequacy of the Assessment's Items

The following four recommendations result from our own analysis of the items, as well as from the reactions of educators we invited to examine sample items provided in the *Framework* and the field test items. Although the reactions to the items was in large part favorable, our recommendations are based on those criticisms that were raised by the people we spoke to.

1. The inclusion of open-ended questions has wide approval. However, the *number* of open-ended questions should be reduced to allow students more time to construct thoughtful responses.

2. There should be other response modes. For example, it would make sense to have students actually perform a task when testing their ability to "read to perform a task."

3. Items designed to assess the ability to read to perform a task should be included at the fourth grade as well as at eighth and twelfth grades.

4. There should be an equal distribution of items for each reading situation at each grade level.

Recommendations about Scoring the Assessment's Items

Throughout our investigations concerns about scoring came up repeatedly. Our four recommendations about scoring are in response to these concerns. There was general agreement that the true measure of this assessment's success lies in the validity of the scoring procedures.

1. Items intended to assess "demonstrating a critical stance" should be reconsidered to determine whether they adequately represent the operational definition provided in the *Framework*.

2. To compensate for the possibility that primary trait scoring of open-ended items may not provide a fair assessment of the performance of students from diverse backgrounds, a

qualitative analysis should be conducted and reported along with the results of the primary trait scoring of responses.

3. The scoring standards or anchor points should be derived from the student responses and not rely heavily on the classifications assigned to items by the item writers.

4. A procedure for independent access to the student responses should be developed so that interpretations from diverse perspectives can be incorporated. This could be done by publishing a small sample of the responses or by instituting a process for larger scale access with additional findings.

Recommendations about Reporting the Results of the Assessment

Many people with whom we consulted felt that special attention should be paid to the manner in which the results of this assessment are reported. The following five recommendations should be carefully considered by those charged with providing information about student performance on this assessment.

1. Special attention should be devoted to the items and scoring rubrics for personal response items to accord with the fact that appropriate personal responses may vary widely and in unpredictable ways. Other formats, such as open-ended interviews, may be more valid for assessing personal response.

2. Many examples of student responses should be displayed in the report documents.

3. While the assessment represents a number of sound ideas about reading, there should be a clear recommendation that teachers not use the assessment as a direct guide for instruction. On the other hand, it could be used to develop instructional objectives.

4. Reporting of the results should emphasize that while every effort was made to broaden the concept of assessment in line with research finding on reading, no large-scale assessment can completely accord with all of the research guidelines. For example, the "reading situation" may approximate "reading for literary experience," but is ultimately still reading for a test; "open-ended responses" are only somewhat open-ended if there are tight time constraints; and so on.

5. Field test information from the special studies should be carefully studied to determine whether the studies succeed in meeting the original intentions. Caution should be taken to avoid making inferences about the data that may not pertain to the original intent of these special studies.

The Special Studies

Several special studies were organized around the 1992 NAEP in Reading. The Integrated Reading Performance Assessment contains two studies, the Oral Reading and Response Study and the Reading Portfolio Components Study. In addition, pilot studies were conducted to investigate the effectiveness of a special study of the metacognitive strategies students use to monitor their reading comprehension. These studies were conducted with a small sample of the students in the 1992 NAEP in Reading. One purpose of these special studies is to explore the feasibility of some new approaches to assessment, another is to get information about some important aspects of reading not easily measured in a large-scale assessment. We briefly discuss each of these studies below.

The Integrated Reading Performance Assessment

The two studies in the Integrated Reading Performance Assessment included only a small sample of students. Each study involved a student being interviewed by a NAEP examiner. The responses of the students in these studies were tape recorded. It was estimated that it would take 45 to 50 minutes of student and examiner time to complete the two studies.

The Oral Reading and Response Study. This study has two facets. The first is the examination of students' reading fluency by timing and analyzing their oral reading. Students are asked to read and respond to a passage that they have already read silently, and from that sample, according to the *Framework*, "An analysis will be made of their oral reading fluency by looking for evidence of the use of phonics, sight vocabulary, semantics, and syntax."

The second facet is a comparison of spoken and written responses to the same assessment items. The students are asked to read aloud and respond to questions about a passage that they have read silently and responded to as part of the regular assessment. Their two response modes--written and spoken--are be compared. These comparisons permit a consideration of the degree to which performance on the open-ended written questions can be affected by students' writing ability.

A number of people have praised the special studies as useful research endeavors for NAEP to undertake. One panel member we interviewed suggested that the results of these efforts should lead to valuable information for future developers of assessments. And that the developers, for example, might find that the oral reading component wasn't necessary--and that that would be useful information.

We take advantage of this comment about oral reading to introduce the most controversial of the topics within the group of special studies. From talks we have had with various panel members, it is evident that the decision to include oral reading in the assessment was not made without a great deal of discussion. This often controversial discussion reflected one of the classic issues in beginning reading instruction--should the emphasis of instruction be on exact word reading or on meaning making? The concern of some was that a measure of whether the students can read the words is very important, the concern of others was that such an activity takes the focus away from reading as meaning making. Nevertheless, the decision was made to include an oral reading component in one of the special studies.

Given this decision, one of our concerns about this special study was the confusion about its content across the three relevant documents. For example, the *Framework* says that students are to be asked to read aloud from a passage they have already read silently as part of the assessment. The *Specifications* say that the students are to read two passages taken from the main assessment, one narrative and one expository (p. 39). The ETS materials for the 1990 field test say that the students are to first read aloud from a book they bring to the session, and then read aloud from a literary passage from the assessment "for five minutes or up to approximately 300 words, whichever comes first" (p. 10).

How the taped oral readings were to be scored is also confusing across the three documents. The *Framework* says that "an analysis will be made of their (the students) oral reading fluency by looking for evidence of the use of phonics, sight vocabulary, semantics, and syntax." The *Specifications* say that the score for fluency "will be based on looking at the number of miscues and the total time taken by the respondent to read the passage" (p. 39). The ETS materials say the administrators will "code a series of miscues" (p. 10).

Given the controversial nature of this segment of the special study, and the decision to gather information about reading fluency, it seemed of particular importance that the goals of this portion of this special study be carefully defined and the procedures to gather data relevant to these goals be carefully evaluated in the field tryouts. Concerns about the reading fluency evaluation were voiced by

BEST COPY AVAILABLE

several of the people we interviewed. One planning committee member felt that the committee's definition of fluency needed to be clarified, and that a better procedure would have been to start off by deciding what data were wanted and then designing a protocol to get those data. On the other hand, this same committee member concluded that "The study is well intentioned and the final product livable . . . ." It is important to emphasize, however, that several people spoke to us about their worry that this segment of the special studies would present an inappropriate model of reading instruction to teachers.

The Reading Portfolio Components Study. According to the *Framework*, the main purposes of the Portfolio Components Study was to gather examples of classroom work in reading, and to find out how students respond to longer texts. Students were asked to bring samples of their classroom work with them, and to discuss them with the examiner, along with self-selected books they had been reading. It was anticipated that these interviews would provide information about the content of classroom reading instruction, along with information about students' self-selected reading, and an opportunity to compare how they discuss these books with their responses to the passages on the assessment. The *Framework* also points out that the Portfolio Components Study will "employ an approach to assessment that is rapidly gaining support in states and districts throughout the country."

Descriptions of the portfolio study are also somewhat different in each of the three relevant documents. For example, the *Specifications* ask that students in the sample be given some well-known books to read two weeks before the special study, the *Framework* and the ETS materials say that each student should bring any single book to the interview.

Many of the respondents to the survey questionnaire liked the idea of a portfolio component of the assessment and many of the people we spoke to praised this aspect of the Integrated Reading Performance Assessment. On the other hand, there were some doubts. The following two comments from interviews we conducted express some of them:

> I have some doubts about the feasibility of the portfolio approach for a national assessment.
>
> I am hesitant about doing portfolio assessment at a national level. If they do this then they should mandate the type of information to be included so that there will be consistency in the sample of work included. I'm concerned about the time and money this type of assessment would cost at a national level. Is this the best way to spend our money?

And a more political doubt is expressed by another:

> I wonder if there is a real interest in doing these studies by NAEP or were they included to appease certain constituencies. I'm not convinced that the special studies represent a serious endeavor. They may simply be political decisions. I do think it is important for NAEP to try and keep everyone satisfied so that the results of NAEP don't get too mired in controversy over the legitimacy of the test.

Others questioned that "portfolio" was the appropriate label for this aspect of the Integrated Reading Performance Assessment, pointing to the limited amount of information that would be gained from this study--based on what students bring in to one interview session--as compared to that of the much more elaborate and wide-ranging classroom based portfolio assessments being developed in many school districts.

The Metacognitive Study

This study was to be piloted in several locations throughout the country. The intent was that, by interviewing students as they read passages, information would be obtained about their awareness of their own comprehension and their use of metacognitive reading strategies.

Four Questions

In addition to the analyses discussed in the preceding sections of this report, we were asked to consider four specific questions. The first question has to do with the field of reading, the second with the nature of the students in the nation's schools, and final two questions with the assessment itself. We will consider each in turn.

1. How can the assessment results best be presented to the professionals in the field of reading given the fact that there are no clearly defined and agreed-upon guidelines for the teaching of reading?

We have several suggestions about the presentation of the data from the 1992 NAEP in Reading to reading professionals.

The presentation of information about NAEP in Reading to the field should be taken very seriously and therefore carefully planned. Information should be made available, far in advance of the announcement of the assessment results, that will permit a discussion of the assessment, the interpretation of its results, as well as their implications for reading instruction.

One of the people we interviewed made the following suggestion:

> NAEP should be more concerned with the use and impact of assessment now that it is becoming a high stakes test. On one hand NAEP doesn't think enough about the impact of the test and on the other they may overestimate the importance of the test. They need to remember the limitations given sampling and limited participation.

All of the documents associated with the NAEP in Reading that are presented to the public should be clearly and carefully written. Our panel members stressed the importance of clear public documents for at least three reasons: the divisiveness of the field, the significant change in the assessment itself, and the importance of state by state reporting. Panel members read both the survey questionnaire and the *Framework* before our two day meeting. Several of them commented that the survey questionnaire offered a clear--if limited--representation of the ideas of the *Framework*, whereas the *Framework* itself was somewhat confusing.

Given the diversity of opinion in the field about reading instruction, it seems extremely important to present the rationale of the assessment, as well as reports of the findings as clearly as possible to avoid unnecessary confusion about what was being assessed, how the assessment was scored, and how the results are interpreted. One panel member urged that reports on student performance include many more examples of student responses than have appeared in previous NAEP reports. Such examples would be especially helpful in illustrating the various anchor points that are established to report different levels of reading performance. It makes perfect sense that reports of performance-based assessments include a generous number of examples to illustrate student performance.

Another panel member urged that a wide range of reading professionals be involved in making decisions about scoring. Earlier in this report we urged that reading professionals be involved in all of the phases

of the assessment. Having groups of reading professionals from differing perspectives participate in establishing scoring and reporting procedures would be an example of such involvement.

2. Does the proposed assessment adequately address the issues of linguistic diversity and varying background knowledge of a multicultural student population?

We have arrived at no clear answer to this question. As is evident, a number of people were concerned about this issue. One of our panel members said that it would be useful and important for the performance of different ethnic groups on the assessment to be examined. Given that much of the assessment involves written responses to open-ended questions it would be helpful to educators to have rich descriptive information about whether there are characteristic responses specific to Latino, African-American, Native-American, Asian-American, or Anglo-American students.

3. Given that it is common practice to adjust teaching so that students will do well on tests, how will student performance be affected by the implementation of the assessment?

In general, most of the people we spoke with thought that the NAEP had the potential to be a positive influence on reading instruction. Most of the survey respondents thought that NAEP should attempt to develop an assessment that can serve as a useful guide to instruction. The members of the NAEP Planning Committee we interviewed were very supportive of this idea. One of them commented:

> The assessment provides a good model for reading instruction. It represents an interactive view of reading. The assessment is far ahead of the schools I work with. The assessment could have a positive impact on instruction. It provides a good instructional model.

There were some alternative perspectives on this issue. One member of our panel said:

> I think assessment guidelines and instructional guidelines are often different and should be different.

Some of the people we interviewed believed it misleading to expect the goals of teaching and testing to line up. At a general level, the 1992 NAEP in Reading accorded well with majority views about teaching. But as we examined more specific aspects such as item distributions, question types, and length of time for responses, there were increasing concerns about its appropriateness as a guide for instruction. Many people argued that no test, however well designed, should serve that role.

4. How will the results of the assessment be explained to the public and policy makers given the possibility that large numbers of students may do poorly on it?

It is our impression that the participants in the colloquia and the members of the panel thought the assessment might be quite difficult for large numbers of students. One colloquia participant commented on the preponderance of high-level questions and the absence of items that assess literal understanding.

> All these questions seem to be meta-questions. Where are the questions of yesteryear that said, "Did Sally buy a red bicycle?" That kind of question doesn't appear. The closest to that kind of question would have to be initial understanding, but for that they've got "What does the author think about this topic?" That's pretty far removed from the color of the bicycle.

The particular aspect that people felt made the assessment more difficult than previous NAEPs was the amount of writing required. People were concerned about the performance record for those students who can read but do not write very well. It was noted that second language learners can fall into this

category. The assessment would underestimate the reading performance of students who are not used to writing in response to what they read. One participant summarized the views of many when she said:

> I am very concerned about measuring reading through writing. I think that is going to be very misleading because anyone who has been having trouble with writing is going to score low on reading.

Conclusions

Our conclusions are primarily in the form of the recommendations we have presented in the body of the report. We focus in this section on some general observations about the efforts that have led to the 1992 NAEP in Reading.

One of the effects of the commendable effort of the Reading Consensus Project to involve as many as possible of the stake-holders in the process was the establishment of checks and balances to deal with the many divisive issues that were of great importance to people in a contentious field. Their plan included allowing for the expression of divergent views so that no single person, group, or institution would seem to be in complete charge of the process or able to claim full responsibility for the products of the process. We believe this plan achieved its goal, and that no single person, group, or institution can claim the NAEP for Reading as its own.

This achievement also has a down side--and that is that there is no *center* to NAEP. While it is an NCES project, NAGB has oversight and administrative responsibilities, and details of the implementation are carried out by diverse groups, agencies, and offices. Critical decisions about what is to be assessed and how assessment is to be implemented are then made in different quarters. It seems that no person or group, even within NAGB is thus in a position to justify the assessment and the articulation of consensus, *Framework*, *Specifications*, items, scoring, and reporting. Given the importance of the 1992 NAEP for Reading, and the importance of the trial state assessment, we recommend that a more evident center of NAEP should provide information about the present assessment, and be the organizational center for the subsequent development of NAEP in the future.

But none of these problems should detract from the achievements of the group of people who worked on this assessment. The 1992 NAEP for Reading is an advance in large-scale assessment. We praise its planners and developers for their achievements. They have moved away from the limited and constrained formats of previous assessments to an assessment that strives to be representative of a contemporary view of reading. The following comment by a member of the Planning Committee helped remind us that the new NAEP should be judged in comparison with other assessment efforts.

> I've been analyzing tests and performance assessments from around the country as part of the CREST grant at the University of Colorado and UCLA and I think this is the best reading assessment that I have seen. I think there are some outstanding questions because they fit well with constructing meaning and elaborating and responding critically.

Finally, we acknowledge that state-by-state reporting may be the source of much of the criticism leveled at the 1992 assessment. This is an important issue that deserves a thoughtful critique. Although it can be considered independently from the quality of the assessment itself, it is important to understand NAEP within the broader context of how assessment is used and viewed by the public at large and educators throughout the United States.

References

Butzow, C. M., & Butzow, J. W. (1989). *Science through children's literature: An integrated approach*. Englewood, CO: Teachers Ideas Press.

Commeyras, M., Osborn, J., & Bruce, B. (in press). *What do classroom teachers think about the 1992 NAEP in Reading?* (Tech. Rep.). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Commeyras, M., Osborn, J., & Bruce, B. (in press). Reading educators' reactions to the reading framework for the 1992 NAEP. *Yearbook of the National Reading Conference*.

Council of Chief State School Officers (1990a). *Report of the consensus process: 1992 NAEP Reading Assessment, NAEP Reading Consensus Project*. Washington, DC: Author.

Council of Chief State School Officers (1990b). *Reading framework for the 1992 National Assessment of Educational Progress*. Washington, DC: Author.

Fish, S. (1980). *Is there a text in this class? The authority of interpretive communities*. Cambridge: Harvard University Press.

Hartman, D. (1990). *Eight readers reading: The intertextual links of able readers using multiple passages*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

International Reading Association (1991). *Assessment resolutions*. Newark, DE: International Reading Association.

Langer, J. A. (1989). *The process of understanding literature* (Tech. Rep. No. 2.1). Albany, NY: State University of New York, Center for the Learning and Teaching of Literature.

Langer, J. A. (1990). The process of understanding: Reading for literary and informative purposes. *Research in the Teaching of English, 24,* 229-257.

Lucas, C. K. (1988a). Toward ecological evaluation. *The Quarterly of the National Writing Project and the Center for the Study of Writing, 10*(1), 1-3, 12-17.

Lucas, C. K. (1988b). Toward ecological evaluation: Part 2. Recontextualizing literacy assessment. *The Quarterly of the National Writing Project and the Center for the Study of Writing, 10*(2), 4-10.

Rosenblatt, L. (1978). *The reader, the text, the poem*. Carbondale: Southern Illinois University Press.

Tompkins, J. P. (1980) (Ed.) *Reader-response from formalism to post-structuralism criticism*. Baltimore: The John Hopkins University Press.

Winograd, P. (1988). Strategic difficulties in summarizing texts. *Reading Research Quarterly, 19,* 404-425.

Author Notes

We are grateful to Edward Roeber and Isabel Beck, members of the National Academy Panel on the Evaluation of the Trial State Assessment Project, and George Bohrnstedt of the American Institute for Research for their astute comments on an earlier version of this report. In addition, we benefitted enormously from conversations with our colleagues at the Center, with faculty of the College of Education at the University of Illinois, and with members of the panel we convened. We interviewed many people who had been involved with the development of the *Framework*. We thank all of them for their observations and for their time. We also wish to thank the people who took the time to fill out the survey, often with extended responses. But, we apologize to those people we could and should have interviewed, but didn't because of lack of time. As the time we were given to accomplish this project marched along, we experienced the tight time constraints that were such conspicuous features of the NAEP Reading Consensus Project.

We also acknowledge the support provided by the National Academy of Education.

Table 1

Questions for NAEP Interviews

Begin by explaining what we're doing and why we're asking for their help. Then ask if they have any questions before beginning with our own questions.

1. What was the extent and nature of your involvement on the Planning Project Committee for the 1992 NAEP in Reading?

2. Can you discuss some of the major issues that were addressed in the sessions you attended?

3. Were you satisfied with the process used to develop the *Framework*?

4. What is your view of the *Framework*? What are it's strengths (weaknesses)?

5. Do you think the *Framework* adequately represents current theories and practices in the field of reading education?

6. Do you think the *Framework* represents a consensus of the field?

7. Can you provide us with any insights regarding the special studies: The Integrated Reading Performance Record (the oral reading study, the reading portfolio, background practices, reading strategies)?

8. Approximately 40% of the items call for written responses. Do you think this places the appropriate emphasis on open-ended responses? Should there be more or less?

9. Do you think the 1992 NAEP provides a good model for reading instruction should anyone decide to teach to the test?

Use all questions for Planning Project. Use 1, 3, 4, 10 for NAGB. Rephrase 1 to focus on NAGB's relations with the planning committee.

Table 2

Percentage of Agreement with ETS on Categorization of 10 Item Blocks

	n (items)	Percentage of Agreement
Initial Understanding	10	90
Developing Interpretation	52	83
Personal Response	25	80
Critical Stance	35	11
Total	122	62

Table 3

Item Distribution Across Reading Situations and Grade Levels

Grade level	Reading for literary experience	Reading to be informed	Reading to perform a task
4	55%	45%	0%
8	40%	40%	20%
12	35%	45%	20%

Table 4

Percentage of Agreement between Panel of Experts and ETS on Item Categorization for Three Blocks from the 1992 NAEP in Reading

	n (items)	Mean	Range
Initial Understanding	4	53	30 - 73
Developing Interpretation	8	57	18 - 90
Personal Response	10	50	10 - 88
Critical Stance	12	23	0 - 60
All 4 Aspects of Reading	34	42	0 - 90
Subcategories (Literary, Informational, Documents)	34	41	0 - 100

Appendix A

SURVEY COVER LETTER

May 1, 1991

Dear Educator:

We are writing to ask for your help as we prepare a review of the 1992 National Assessment of Educational Progress (NAEP) in Reading. The reading NAEP for 1992 will be based on the ideas and recommendations set forth in a document called the **Reading Framework**. The National Academy of Education has asked us to examine this document to determine the extent to which it represents a consensus about reading among professionals in the field of education.

The review we prepare will be one of ten such reviews of various aspects of the 1992 NAEP that have been commissioned by the National Academy. These reviews will form the basis of a report that will be presented to Congress in October.

So that the ideas and recommendations of the **Reading Framework** could be commented on by a number of educators, we have prepared a survey questionnaire based on the content of the **Framework**. We hope very much that you will find the survey of interest, and that you will complete it, and return it to us. Most questions can be answered by circling a single code number following the question. We encourage you to comment on any or all of the items that are of particular importance to you.

To provide some background information, we have enclosed a brief overview that describes the NAEP, the development of the **Reading Framework**, and some special features of the 1992 reading assessment. If you have any questions, please call us at 217-333-6551.

We would be most appreciative if you could return the survey in the envelope provided by May 31, or otherwise, at your earliest convenience.

Your responses and comments will be invaluable to us as we prepare our review for the National Academy. Thank you very much for your time and your effort.

Best wishes,

Bertram Bruce
Professor of Education
University of Illinois

Jean Osborn
Associate Director
Center for the Study of Reading

Michelle Commeyras
Project Associate
Center for the Study of Reading

Janet Salm
Project Assistant
Center for the Study of Reading

Appendix B

OVERVIEW

What Is The National Assessment of Educational Progress?

The National Assessment of Educational Progress -- "the Nation's Report Card" -- is mandated by Congress. Every two years, NAEP assesses the performance of more than 120,000 fourth-, eighth-, and twelfth-grade students in the nation's schools. The purpose is to gather information about students' performance and about changes in their performance over time.

Since 1969, NAEP has conducted seven assessments in reading, six each in science and mathematics, five in writing, two each in music and art, and one in computer science. NAEP has also conducted special assessments in citizenship, U. S. government, U. S. history, literature, social studies, and other areas.

The National Assessment Governing Board decides on the subject areas to be assessed, including those specified by the Congress. It also is responsible for identifying appropriate achievement goals for each age level; for developing the objectives, specifications, and procedures for each test; for setting the data analyzing and reporting guidelines; and for determining procedures for interstate, regional, and national comparisons based on the data.

The Reading Framework for the 1992 NAEP

In 1989, the National Assessment Governing Board contracted with the Council of Chief State School Officers to develop a rationale and give recommendations for the 1992 National Assessment of Educational Progress for Reading. Because of the diverse and often conflicting opinions about reading and its assessment held by reading educators and others in the field of education, the Council created the Reading Consensus Project and charged it with developing an assessment framework that would be acceptable to the field as a whole.

In response to this charge, the Reading Consensus Project appointed committees composed of teachers and administrators, members of state departments of education, university professors whose specialties included reading and assessment, and representatives from a number of educational, business, and professional organizations. These committees met between October 1989 and February 1990.

The members of these committees were dedicated not only to developing a framework that would reflect the consensus of the field of reading but also to ensuring that the framework would be consistent with sound, contemporary research about reading. To this end, drafts of the developing framework were sent out for comment to a large number of chief state school officers; state assessment directors; school administrators and teachers; professors of reading, education, and psychology; and assessment experts. The committees' final version of the Framework was submitted to the National Assessment Governing Board in June 1990.

BEST COPY AVAILABLE

Changes from Earlier NAEP Assessments

The Reading Framework proposes some major changes in the 1992 NAEP assessment:

1. Authentic Texts. The assessment will use reading passages drawn from books and articles like those students read in school and on their own, rather than passages written solely for testing purposes, such as for assessing particular reading skills. These passages will be much longer than those used in previous NAEP assessments. The eighth-grade students will, for example, read an entire short story, a newspaper article, and a complete set of instructions.

2. Three Reading Situations. The passages students will read are classified into three types of reading situations: (1) reading for literary experience, (2) reading to acquire information, and (3) reading to perform a task. Reading for literary experience will be assessed by having students respond to questions about a short story or a poem. Reading to acquire information will as assessed by having them respond to questions about a newspaper article or a textbook selection. Reading to perform a task will be assessed by having them respond to questions about an instruction manual or a train schedule.

3. Assessment of the Cognitive Aspects of Reading. The Framework recognizes that proficient readers use a range of cognitive abilities to construct meaning from a text and to elaborate upon and respond critically to it. The Framework also recognizes that these cognitive aspects of reading are not sequential or hierarchical and do not represent a set of subskills. The Framework proposes that these cognitive aspects be assessed within each of the three reading situations.

The ability to construct meaning, for example, will be assessed by two types of questions:

> **Forming initial understanding questions,** which require students to provide an initial impression or global understanding of what they have read.
>
> **Developing an interpretation questions,** which require students to go beyond their initial impressions to create a more complete understanding of what they have read.

The ability to elaborate on or respond critically to a text will also be assessed by two types of questions:

> **Personal reflection and response questions,** which require students to connect knowledge from the text with their own personal background knowledge.
>
> **Demonstrating critical stance questions,** which require students to consider a text objectively.

4. Multiple-Choice and Open-Ended Questions. Approximately 60% of assessment time will be spent on multiple-choice questions, 40% on open-ended questions. Some of the open-ended questions will be designed for one- or two-sentence answers, others for more extended written responses. Primary-trait scoring will be used for the extended responses, and scoring rubrics will be created for each question.

5. **Two Special Studies: Integrated Reading Performance Assessment**

Several types of information about the reading performance of students will be collected from special studies with subsamples of students.

> Oral Reading. Tape recorded interviews will be used to examine the oral reading fluency of fourth-grade students.
>
> Portfolio Assessment. Taped interviews will also be used to gather information about classroom reading instruction. For these portfolio type activities, the students will talk about both their independent and classroom reading assignments. In addition, they will be asked to bring samples of their written work to the interview.

6. State-by-State Reporting. In the past, NAEP considered the nation's students as a single body, and reported its data on the basis of grade level, gender, ethnicity and type of community (rural or urban). In response to requests from both state and national educational leaders, NAEP data now will also be reported by state. The 1990 NAEP mathematics assessment will provide state-by-state information, and the 1992 fourth-grade reading assessment will do the same. It is anticipated that in the future, all NAEP data will be reported as both a national and as state-by-state assessments.

Appendix C

RESULTS OF THE SURVEY BASED ON THE CONTENT OF THE READING FRAMEWORK FOR THE 1992 NAEP

A. Characteristics of Proficient Readers

1. The Reading Framework identifies characteristics of proficient, or good, readers. To what extent do you agree or disagree with each of the following statements about these characteristics.

	Strongly agree	Agree	Disagree	Strongly disagree
a. Good readers possess the knowledge, behavior, and attitudes that allow for continual learning through reading . .	87.55%	11.62%	0	0.41%
b. Good readers read with enough fluency so that they can focus on the meaning of what they read, rather than devoting a lot of attention to figuring out the words	88.80%	10.37%	0	0
c. Good readers use what they already know to understand the text they are reading	87.97%	12.03%	0	0
d. Good readers extend, elaborate, and critically judge the meaning of what they read	74.27%	23.65%	0.41%	0.41%
e. Good readers plan, manage, and check the progress of their reading .	57.26%	36.10%	3.73%	1.24%
f. Good readers use a variety of effective strategies to aid their understanding	77.59%	19.50%	1.66%	0.41%
g. Good readers can read different types of texts and can read for different purposes	83.40%	14.94%	0	0.41%

B. **Views of Reading**

2. The Reading Framework makes the following statements about reading. Indicate the extent to which you find each definition or statement acceptable.

	Very acceptable	Acceptable	Unacceptable	Very unacceptable
a. Reading is a complex process that involves an interaction among the reader, the text, and the context, or situation . . . . . No Response = 0.41%	85.06%	12.45%	1.24%	0.83%
b. The term "reading literacy" connotes more than basic or functional literacy. Specifically, it connotes knowing when to read, how to read, and how to reflect on what is being read . . . . . No Response = 0.83%	69.29%	23.24%	4.56%	2.07%
c. Proficient reading is essential for successful functioning in schools, homes, and workplaces . . . . . No Response = 0.83%	70.54%	24.90%	2.90%	0.83%
d. Proficient reading contributes to a sense of personal satisfaction . . . . . No Response = 1.24%	67.63%	26.97%	3.32%	0.83%

C. **Reading Situations**

3. In the NAEP, students' reading ability will be assessed in three situations: reading for literary experience, reading to be informed, and reading to perform a task. In your opinion, how important is it to assess students' reading ability in each situation.

	Absolutely essential	Very important	Moderately important	Somewhat important	Not at all important
a. Reading for literary experience (short stories, poems, essays) . . . . . No Response = 2.07%	42.74%	41.08%	12.86%	0.41%	0.83%
b. Reading to be informed (magazine and newspaper articles, encyclopedias, textbook chapters) . . . . . No Response = 1.66%	76.76%	19.09%	2.49%	0	0
c. Reading to perform a task (bus and train schedules, directions for games, recipes, maps, etc.) . . . . . No Response = 1.24%	78.42%	14.52%	3.73%	1.66%	0.41%

BEST COPY AVAILABLE

4a. The following table shows the proportion of items in the NAEP allocated at each grade level to each type of reading situation. This distribution of items is intended to reflect the changing demands made of students as they progress through school.

Grade level	Reading for literary experience	Reading to be informed	Reading to perform a task
4	55%	45%	0%
8	40%	40%	20%
12	35%	45%	20%

Do you agree with this NAEP allocation of items?

Yes 31.12%
No 59.34%

No Response 9.13%

4b. If you disagree, use the table below to show how you would reallocate the proportion of items.

Grade level	Reading for literary experience	Reading to be informed	Reading to perform a task
4	48.93 (9.22)	41.34 (7.57)	9.39 (9.68)
8	39.08 (5.40)	40.14 (4.46)	20.83 (4.98)
12	34.84 (6.26)	43.80 (5.66)	21.41 (5.81)

D. **Cognitive Aspects of Reading**

5. The Framework identifies four cognitive aspects of reading to be assessed within each of the three reading situations. Each of these aspects is described below. Indicate how important you believe each to be.

	Absolutely essential	Very important	Moderately important	Somewhat important	Not at all important
a. Forming an initial understanding requires the reader to provide an initial impression or global understanding of what was read . . .	56.43%	35.27%	3.73%	2.49%	0.41%
No Response = 1.66%					
b. Developing an interpretation requires the reader to develop a more complete understanding of what was read by linking information across parts of a text as well as by focusing on specific information in the text	67.22%	27.80%	2.07%	1.66%	0
No Response = 1.24%					
c. Personal reflection and response requires the reader to connect knowledge from the text with his or her own personal background knowledge	68.88%	23.65%	4.15%	1.66%	0.83%
No Response = 0/83%					
d. Demonstrating a critical stance requires the reader to consider the text objectively and involves a range of tasks including critical evaluation, comparing and contrasting, application to practical tasks, and understanding the impact of such text features as irony, humor, and organization	53.53%7	34.44%	8.30%	0.83%	0.83%
No Response = 2.07%					

6a. In your opinion, are there other cognitive processes of reading that need to be represented in the Framework?

Yes	19.50%
No	68.88%
No Response	11.62%

E. **Open-ended Items**

BEST COPY AVAILABLE

7a. Approximately 40% of the assessment time will be spent on open-ended items. This is substantially more time than has been given to such items in any previous NAEP assessment. The following rationale is given for the increased use of open-ended items is that they provide a means for examining whether readers can generate organized, carefully thought-out responses to reading. Also, open-ended items more closely resemble the real world tasks that students must perform outside of school.

Do you support the inclusion of open-ended items in the NAEP Assessment?

Yes	95.44%
No	3.32%
No Response	1.24%

7c. Do you agree with the rationale given above for including more open-ended items?

Yes	86.72%
No	5.81%
No Response	7.47%

7d. Do you think the proportion of open-ended items (40%) in the NAEP is:

Too much	12.45%
The right amount	62.24%
Too little	0
No Response	8.71%

F. **Passage Selection**

8a. The passages used in the NAEP will be authentic, full-length texts that students are likely to encounter in everyday reading (i.e., short stories, newspaper articles, bus schedules, textbook chapters, pages from telephone directories). The passages will not be paragraphs written solely to assess specific reading skills.
In your opinion, should the assessment use . . .

Only authentic passages,	66.39%
Only passages written to test specific skills, or	1.24%
A combination of authentic passages and passages written to test specific skills?	30.29%
No Response	2.07%

G. **Teaching to the Test**

9. Although the Framework claims that assessment should not drive instruction, it also states that the NAEP assessment must be an appropriate guide to instruction.

   a. Do you believe that NAEP should attempt to develop an assessment that can serve as a useful guide to instruction?

Yes	72.20%
No	25.73%
No Response	2.07%

H. **Special Studies**

10. Several types of information about the reading performance of students will be collected from special studies with small subsamples of students. These studies are listed below. To what extent do you believe that each of these studies is needed?

	Not at all 1	2	3	4	To a very great extent 5
a. Oral Reading. Fluency will be assessed by timing and analyzing students' oral reading . . . . . No Response = 1.24%	7.05%	20.75%	34.44%	21.58%	14.94%
b. Portfolio Assessment. Portfolio activities will be used to gather and analyze examples of actual classroom work in reading as well as to gather information about what the students read in class and on their own . . . . . No Response = 2.07%	2.49%	2.07%	8.71%	28.63%	56.02%
c. The Metacognitive Study. Readers' awareness of their own comprehension and their use of effective reading strategies will be assessed, analyzed and reported as descriptive data . No Response = 1.66	2.90%	2.07%	12.03%	29.46%	51.87%

I. Goals of the 1992 NAEP in Reading

11. The committees that developed the Framework were given the following set of guidelines. Indicate how well you believe the assessment will meet each one of the guidelines.

	Not at all well 1	2	3	4	To a very great extent 5
a. To focus on outcomes (is performance oriented), rather than representing an instructional or theoretical approach . . . . . No Response = 6.64%	2.49%	9.13%	24.48%	41.91%	15.35%
b. To address changing literacy needs for employability, personal development, and citizenship . No Response = 5.39%	2.07%	4.56%	27.80%	39.42%	20.75
c. To expand the scope of assessment strategies by including open-ended questions and special studies on oral reading, portfolio assessment, and reading strategies . . . . . No Response = 4.56%	0.41%	1.24%	12.03%	34.44%	47.30%
d. To reflect contemporary research on reading and literacy . . . . . No Response = 5.39%	1.24%	2.49%	16.18%	34.85%	39.83%
e. To provide information for policy makers and educators that will assist in the improvement of educational performance . . . . No Response = 5.81%	4.56%	8.71%	14.94%	41.08%	24.90%

J. State-by-State Reporting

12. The 1992 NAEP for reading will provide state-by-state as well as national reports of student performance. How do you feel about state-by-state reporting of student performance?

Strongly in favor 31.12%

Somewhat in favor . . . 29.88%

Somewhat opposed . . . 15.35%

Strongly opposed 20.33%

No Response 3.32%

K. **Personal Information**

13a. Are you currently employed as a teacher at the elementary, secondary, or college level?

Yes	59.75%
No	38.17%
No Response	2.07%

b. Indicate the grade level(s) you teach. (CIRCLE ALL THAT APPLY.)

K-2	11.62%
3-5	10.78%
6-9	6.64%
10-12	3.73%
Special Education	1.24%
Chapter 1	5.81%
Undergraduate	32.37%
Graduate	34.02%
Other (SPECIFY)	4.98%

_____

14a. Do you currently hold an administrative position at the school, district, or state level?

Yes	36.51%
No	59.34%

b. What administrative position do you hold?

Superintendent of Public Instruction	1.24%
Superintendent/Assistant Superintendent	1.24%
Principal/Assistant Principal	2.90%
Reading Coordinator/Supervisor/Consultant	21.16%
Other (SPECIFY)_____	9.96%

15a. Do you currently hold a college or university position?

Yes	47.72%
No	47.72%

b. What position do you hold?

Professor	36.93%
Administrator	1.24%
Research Associate	2.49%
Other (SPECIFY)	7.05%

16. Indicate the area that best represents your field of specialization.

Reading	78.84%
Writing	1.66%
Assessment	3.32%
Other (SPECIFY)	14.52%

17. How many years have you been employed in the field of education?

0-5 years	3.32%
6-10 years	8.71%
11-15 years	17.43%
16-25 years	46.47%
Over 25 years	23.24%

18. How familiar are you with previous National Assessments of Educational Progress in Reading?

Very familiar	34.85%
Somewhat familiar	49.79%
Not at all familiar	14.11%

Thank you very much for your cooperation.